SNAPS: A TRILOGY PARTS – [0] [1] [2] COMBINED

Shoumen Datta

HAPHAZARD REALITY – IOT IS A METAPHOR

Analyst

Accepted Manuscript

rsc.li/analyst

Post hoc support vector machine learning for biosensors based on weak protein-ligand interactions

Y. Rong[a], A.V. Padron[a], K. J. Hagerty[a], N. Nelson[b], S. Chi[c,d], N. O. Keyhani[d], J. Katz[e], S.P.A. Datta[f,g], C. Gomes[h], and E.S. McLamore[*a]

Impedimetric biosensors for measuring small molecules based on weak/transient interactions between bioreceptor and target analyte are a challenge for detection electronics, particularly in field studies or in analysis of complex matrices. Protein-ligand binding sensors have enormous potential for biosensing, but accuracy in complex solutions is a major challenge. There is a need for simple *post hoc* analytical tools that are not computationally expensive, yet provide near real time feedback on data derived from impedance spectra. Here, we show use of a simple, open source support vector machine learning algorithm for analyzing impedimetric data from multiple protein-based biosensors, and we show that the tool can be used for point of need applications in analysis of small molecules such as acetone using a mobile phone. In all conditions tested, the open source classifier was capable of performing as well, or better, than equivalent circuit analysis for characterizing weak/transient interactions between a model ligand (acetone) and a small chemosensory protein derived from tsetse fly. In addition, the tool has a low computational requirement, facilitating use for mobile acquisition systems such as mobile phone. The protocol is deployed through Jupyter notebook (an open source computing environment available for mobile phone, tablet, or computer use) and the code was written in Python. For each of the applications we provide step-by-step instructions in English, Spanish, Mandarin, and Portuguese to facilitate widespread use. All codes were based on *scikit-learn*, an open source software machine learning library in the Python language, and were processed in Jupyter notebook, an open-source web application for Python. The tool can easily be integrated with mobile biosensor equipment for rapid detection, facilitating use by a broad range of impedimetric biosensor users. This post hoc analysis tool can serve as a launchpad for convergence of nanobiosensors in planetary health monitoring applications based on mobile phone hardware.

Introduction

Biosensors offer rapid analysis of targets ranging from small molecules, to biomolecules or cells, and can be applied across a wide variety of planetary health applications in medical, agricultural, and environmental analysis[1, 2]. With the advent of mobile phone electrochemical and plasmonic acquisition systems [3-5], the portfolio of biosensors used in applied field studies is rapidly expanding. Biosensor accuracy, speed, range, and limit of detection are a function of the nature of molecular interactions between target analyte and bioreceptor structure, the transduction mechanism, inclusion of nanomaterials which enhance transduction, type of detection hardware, and acquisition approach (including *post hoc* analysis).

Among the various transduction approaches, electrochemical biosensors are one of the most popular device types, and most current devices combine electroactive nanomaterials (e.g., graphene, nanometal, electropolymers) with biorecognition structures such as enzymes, antibodies, or aptamers, among others [6-10]. Use of transducer nanomaterials enhances signal acquisition, while the biological material is used to impart selective targeting and in some cases, catalyze a reaction[11-13]. Impedimetric biosensors are most commonly developed based on Faradaic impedance (with redox couple in solution), but label-free biosensors using non-Faradaic impedance (absence of redox couple) are increasing in popularity[14, 15]. In either case, the output impedance depends on changes in the interfacial electron transfer resistance and/or electrostatic repulsion that result from steric hindrance caused by interactions of the target and bioreceptor[16-18].

Interpretation of impedimetric biosensor data is often not trivial, particular for fast electron transfer processes in nanomaterial-modified electrodes, non-Faradaic impedance, or weak/transient interactions between bioreceptor and target. Impedance data are usually fit to an equivalent circuit model using Chi[2] testing, and parameters derived from the model to describe the underlying electrochemistry. Changes in equivalent circuit parameters are commonly reported as sensor output, although impedance at a single frequency is occasionally used as sensor output[14]. Equivalent circuit analysis is based on combinations of the Principle of Superposition, Ohm's Law, and Kirchoff's Laws, and is very accurate for simple electrode geometries with homogenous surfaces. However, as circuit models are assumed *a priori,* there is not necessarily a correspondence between circuit elements and underlying physico-chemical processes[19]. Furthermore, inclusion of transducer nanomaterials on the sensor surface complicates the equivalent circuit model, requiring additional "fitting" elements. Thus, interpreting impedance data in complex solutions or with complex electrode geometries is challenging,

a. Agricultural & Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida.
b. Biological & Agricultural Engineering, North Carolina State University.
c. Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences; Key Laboratory of Microbial Resources, Ministry of Agriculture, Beijing, China.
d. Department of Microbiology and Cell Sciences, Institute of Food and Agricultural Sciences, University of Florida.
e. Department of Oral and Maxillofacial Diagnostic Sciences, University of Florida
f. MIT Auto-ID Labs, Department of Mechanical Engineering, Massachusetts Institute of Technology,
g. Biomedical Engineering Program, Department of Anaesthesiology, Massachusetts General Hospital, Harvard Medical School.
h. Department of Mechanical Engineering, Iowa State University
† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

and is sometimes more art than science[20]. The main challenge for planetary health biosensors is to balance enhancing conductivity with transducer nanomaterials (improving limit of detection) while limiting computational cost (maintaining speed), and at the same time developing simple label-free devices that can be used in diverse applications (ensuring robustness).

To improve limit of detection, many labs coat electrodes with nanomaterials such as graphene and/or nanometal, which is known to significantly enhance conductivity and electroactive surface area[21-23] while significantly decreasing charge transfer resistance (R_{ct}). This results in fast electron transfer processes, where Faradaic current is represented by a near-linear Nyquist plot. In a classic Randles-Ershler equivalent circuit, *post hoc* sensor analysis is usually constrained to R_{ct}, as other circuit parameters are a function of the solution resistance or inductance, which are not strong indicators of molecular interactions between bioreceptor and target analyte. This situation is particularly challenging for weak/transient interactions, where more complex circuit models with fitting parameters are required, increasing the computational cost while producing output parameters that have no physico-chemical meaning in the electrochemical circuit. There is a need for simple *post hoc* analytical techniques that can be used for point of need biosensors, particularly for field applications.

Machine learning has emerged as a powerful tool for analysis of sensor data in a wide range of applications, including: flow cytometry[24], electronic tongue/nose[25-27], wearable sensors[28, 29], whole organism biosensing[30, 31], protein detection[32], sensor material optimization[33], food safety risk analysis[34], environmental pollutant monitoring[35] and multiplexing sensors arrays [36-38]. These techniques provide a platform for development of systems-level planetary health solutions that focus on convergence of nanobiosensors, mobile phone sensor acquisition, and data analytics.

Here, we present an open source machine learning algorithm applied for label-free biosensors based on weak/reversible interactions that can be used with common mobile hardware (mobile phone or tablet). We demonstrate the utility of this approach for analysis of a biosensor based on reversible interactions between a small molecule (acetone) and insect-derived chemosensory proteins. *In vivo*, CSP solubilize volatile odorants and facilitate transport to downstream odorant receptors (ORs) through reversible association/disassociation [39, 40]. This represents a model impedimetric biosensor based on interactions between low molecular weight binding proteins and small molecules. Biosensors based on CSP are becoming popular, but the transient ligand interactions and relationship to underlying electrochemistry are not well documented. As a comparison, we also analyzed weak/transient protein-biomolecule interactions (both protein-DNA and protein-protein) using well-known binding proteins with equivalent circuit analysis and machine learning. The protocol is based on Jupyter notebook (open source computing environment available for mobile phone, tablet, or computer use) and the code was written in Python. For each of the applications we provide step-by-step instructions in English, Spanish,

Mandarin, and Portuguese to facilitate widespread use for a variety of applications. The open source tool can easily be integrated with mobile biosensor equipment for rapid detection, facilitating use by a broad range of biosensor users.

Methodology

Strains and reagents

Escherichia coli strain Rosetta DE3 (Promega, Madison WI, USA) was routinely grown in Luria-Bertani broth (LB) and/or on LB-agar (1.5%) plates containing 50 µg/mL kanamycin. All reagents and chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) or Thermo Fischer Scientific (Waltham, MA, USA) except as noted. Potassium ferrocyanide (K_4FeCN_6), potassium ferricyanide ($K_3[Fe(CN)_6]$), and potassium chloride (KCl) were purchased from EMD chemicals (Billerica, MA, USA). Ni- and Co-NTA agarose was purchased from Gold Biotech (St. Louis, MO, USA). Thrombin was purchased from Amersham-Pharmacia Biotech (Little Chalford, UK). Polycrystalline diamond suspensions (3 and 1 mm) alumina slurry (0.05mm) were purchased from Buehler (Lake Bluff, IL, USA).

Electrochemical analysis

For all electrochemical analysis, a three-electrode system was used together with an electrochemical impedance analyzer (ERZ100, eDAQ, Colorado, USA). All electrochemical impedance spectroscopy (EIS) studies used Pt/Ir working electrodes (MF-2013, 1.6 mm diameter, BASi, West Lafayette, USA), Ag/AgCl reference electrodes (BASi, West Lafayette, USA) and platinum auxiliary electrodes (BASi, West Lafayette, USA) with nanoplatinum deposited as previously described[41, 42]. Before all experiments, the Pt/Ir working electrodes were polished with two sizes of polycrystalline diamond suspensions (3 and 1 µm), rinsed with methanol, polished with alumina slurry (0.05µm) and then rinsed with deionized water. Probes were cleaned in a sonication bath in DI water for 15 min, then with 0.1 M H_2SO_4 using cyclic voltammetry (CV) at a potential range of -1.0V to +1.0V until the peak current changed by less than 1%, and then finally cleaned in a sonication bath in DI water for 15 min. To ensure consistency during adsorption studies, electrodes were fitted with a plastic cap that was 3D printed on a Makerbot Replicator 2 Desktop 3D printer (see supplemental Figure S1 for specifications).

EIS analyses were conducted at 0.25V (DC), with a 100mV (AC) amplitude in the range of 100 kHz to 1 Hz in a solution with 2.5 mM potassium ferrocyanide ($K_4[Fe(CN)_6]$), 2.5 mM potassium ferricyanide ($K_3[Fe(CN)_6]$), and 100 mM potassium chloride (KCl). For equivalent circuit analysis, all EIS data was transformed to Nyquist Plots and analyzed using ZMAN (WonATech, South Korea) 2.2 software or support vector classification analysis as noted.

Protein expression and purification

Recombinant insect chemosensory proteins (CSP) derived from *Glossina morsitans* (Gmm, tsetse fly) were heterologously expressed and purified from *E. coli* hosts using the methods described in detail by Song et al[43]. Briefly, GmmCSP3 sequences were identified from genomic databases, codon optimized for *E. coli* expression, and synthesized/cloned into a

pUC vector (Genewiz, Planefield, NJ). Expression constructs were synthesized with a 10X-histidine (10X C terminal His) tag and transformed into *E. coli* host cells. Single colony isolated via selection for ampicillin resistance on LB/ampicillin plates (100 µg/ml). Cells were harvested by centrifugation, washed with Co^{2+} equilibrium buffer, and suspended in the buffer. Protein purification was achieved using Co^{2+} affinity chromatography and elution of the bound protein with increasing concentrations of imidazole. Purity of the protein was assessed by SDS- polyacrylamide gel electrophoresis (SDS-PAGE) and pure fractions were pooled and dialyzed (3000kDa) against buffer (20 mM HEPES, pH 7.5). Protein concentration was quantified with SDS-PAGE, and samples were frozen at -80 °C until used.

Expression and purification of TATA binding protein (TBP) and multiprotein bridging factor 1 (MBF1) from *Beauveria bassiana* have been described in detail previously [43]. Briefly, the coding sequences for both genes were codon optimized for expression in *E. coli* and synthesized (Genewiz) as above for the CSP proteins. Expression plasmids were transformed into competent *E. coli* Rosetta DE3 cells for expression and purification as above. Purified proteins were dialyzed, aliquoted and stored at -80°C until used. After protein elution using, purity was confirmed by SDS-PAGE. Protein concentration was determined using Pierce™ BCA Protein Assay Kit (Thermo Scientific).

PCR products were digested, and then cloned into respective sites of an expression vector to produce plasmids. Expression plasmids were transformed into competent *E. coli* Rosetta DE3 cells and transformed *E. coli* were cultured in LB broth, harvested, lysed, and then purified using Ni- or Co-NTA agarose columns. Purified 10XHis-tagged proteins were aliquoted and stored at -80°C until used. After protein elution using imidazole buffers, purity was confirmed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and protein concentration was determined using Pierce™ BCA Protein Assay Kit (Thermo Scientific).

Protein adsorption onto sensor surface and sensor characterization

For characterizing acetone-CSP interactions, the optimal CSP concentration from previous studies [44] was used or all testing. Briefly, 2 µL of His-tagged GmmCSP3 (9.23 mg/mL) was drop cast on the surface of an electrodes, dried at 20°C for 5 minutes, and rinsed three times with DI water. A 5mM acetone stock solution was prepared in DI water, which is representative of salivary acetone levels for patients with DKA [45]. Where noted, aliquots (2 µL) of acetone stock solution were drop cast on the surface of the biosensor, stored at 20°C for 2 minutes, and rinsed with DI water three times prior to testing.

For biosensors based on protein-biomolecule interactions, the concentration in each experiment was based on Song et al [43]. A 2 µL aliquot of His-tagged TBP was first drop cast on the surface of the electrode, agitated gently, allowed to dry at 20°C for 5 minutes, and rinsed with DI water prior to impedance analysis. Next, 2 µL aliquots of MBF1 (no His-tag) or TATA[1] (a 40 bp DNA sequence containing two potential TATA motifs) was drop cast onto the TBP-functionalized electrode, dried at 20°C for 5 minutes, and then rinsed with deionized

water three times prior to impedance analysis. Control experiments included using uncoupled (no TBP) surfaces as well as TBP-coupled + bovine serum albumin (BSA) solutions and TBP-coupled + TATA[0] (a 35 bp DNA sequence lacking the TATA sequence in TATA[1]).

Data analysis and statistics

EIS plots were analyzed with ZMAN 2.2 using an equivalent circuit model based on Chi^2 analysis. Equivalent circuit parameters, namely solution resistance (R_s), charge transfer resistance (R_{ct}), Warburg impedance (Z_w), double layer capacitance (C_{dl}), and constant phase element (Q) were estimated using Chi^2 fitting in the ZMAN software.

Nyquist and Bode plots were generated with ZMAN 2.2 and several key values were extracted within the software from equivalent circuit analysis. Namely, the Nyquist with equivalent circuit analysis were used to extract R_s, R_{ct}, Z_w, and C_{dl} from a Randels-Ershler equivalent circuit. Bode plots were used to extract the impedance at a given cutoff frequency and associated phase angle. In addition to the Randels-Ershler circuit, various equivalent circuit models (shown in supplemental S5) were tested with the model search function in ZMAN software where noted.

Support vector machine (SVM) classification

For protein-ligand interactions, EIS data were exported and transformed into samples with 152 features that represent both real and imaginary impedance at frequencies from 100kHz to 1Hz. The number of features was selected to satisfy expected confidence levels for principle components analysis. A total of 54 EIS scans were randomly split into two groups, with 80% of the data used as the training set and 20% used as the testing set. Each of the 54 data sets were binary labeled, with baseline impedance data in the absence of acetone labeled as "0", and labeled 1 in the presence of 5mM acetone. EIS data for both baseline (no acetone) and positive (5mM acetone) experiments were standardized and transformed into a two-dimensional dataset, and then mapped in a new data space. To initially screen the data, the four most common types of SVM kernels [46] were used to screen the data. A shuffled K-fold cross validation was used for all applications of SVM in this study [47]; the training dataset was divided into ten folds and shuffled,, with 20% of the total data used for testing. The test accuracy shown for each kernel is the percentage of the prediction accuracy based on the decision boundaries.

Prior to running the SVM algorithm, principal component analysis (PCA) was applied through singular value decomposition (SVD) to reduce the 152 features to two principal components. PCA was used to reduce the dimension of 152 features in the raw EIS data to a two-dimensional principal components matrix. Depending on the number of components to extract, full or randomized truncated SVD was used; this procedure was performed in LAPACK [48]. To ensure generalizability across other varied application-specific biosensors, code screens were prepared for four types of SVM kernels (linear, sigmoidal, radial basis function, and polynomial) to identify which approach best segregates the

training data. This feature of the open source algorithm allows the user to select the most appropriate kernel for a given analysis by comparing the cross-validation results across kernel types. SVM hyperparameters (C and gamma) were optimized using grid search and random search methods[46, 49]. C is a tradeoff between misclassification and simplicity of the decision surface. Gamma is proportional to the radius of influence for selected support vectors [50]. All SVM codes were produced with "scikit-learn", an open source machine learning library in Python [50], and were processed with Jupyter notebook, an open-source web application for Python (see supplemental section for step-by-step instructions in English, Spanish, Mandarin, and Portuguese and Python code).

To validate the functionality of the SVM classifier for a well-known detection system, a TBP-protein and TBP-DNA biosensor were fabricated based on published methods[43]. The biosensor is based on interactions between TBP and either multiprotein bridging factor 1 (MBF1) (protein-protein binding) or TBP and TATA (protein-DNA). His-tagged TBP was first adsorbed to the electrode surface, and then EIS was used to study the interactions of TBP with either MBF (a 17 kDa protein) or TATA[1] (a TBP-binding 40mer DNA sequence). As a control, EIS data were also recorded after addition of buffer, a non-binding protein (BSA), and a non-binding 35mer sequence (TATA[0]). To challenge the approach for detection of small molecules, **a** CSP biosensor for detecting acetone was **also** developed. The experimental conditions were based on levels relevant to diagnosis of diabetic ketoacidosis (DKA), a potentially fatal outcome from complications associated with diabetes.

All experiments were repeated in triplicate, resulting in a total of 54 data sets. Analysis of variance (one-way ANOVA with Games-Howell method and 99% confidence) and student's t-test (two-sample t-test with 99.9% confidence) were performed for analyzing EIS data derived from equivalent circuit modeling as noted. All error bars represent the standard deviation of the arithmetic mean.

Results & Discussion

First, the functionality of the SVM classifier was validated for a well-known detection system using TBP-protein and TBP-DNA based on Song et al[43]. This well-documented biosensor produces large changes in impedance after target binding, and serves as a simple case study for the machine learning tool. The biosensor is based on interactions between TBP and either MBF (protein-protein binding) or TBP and TATA[1], a 40 mer nucleotide sequence containing the TATA motif that is the recognition sequence bound by TBP, (protein-DNA). His-tagged TBP was first adsorbed to the electrode surface, and then EIS was used to study the interactions of TBP with either MBF (a 17 kDa protein) or TATA[1]. As a control, EIS data were also recorded after addition of buffer, a non-binding protein (BSA), and a 35-mer nucleotide sequence lacking the TATA motif (TATA[0]).

Representative Nyquist plots show that adsorption of His-tagged TBP on the sensor surface caused a significant increase in R_{ct}, as expected (**Fig 1a**). Binding between TBP and MBF also

resulted in a significant change in charge transfer resistance, as did binding between TBP and TATA[1]. A Randles-Ershler equivalent circuit (Chi[2]=1087 ± 212) was used to extract R_s, R_{ct}, Z_w, C_{dl} for each experiment (see supplemental Table S1 for details). Similar to other manuscripts in the literature [51], R_{ct} was used as the most accurate parameter for characterizing protein-biomolecule interactions (**Fig 1b**). For comparison, addition of BSA or buffer did not result in any significant change in impedance due to non-specific binding (see supplemental Figure S2). EIS data was further analyzed by SVM classification by dividing the dataset into groups of TBP, TBP+MBF, and TBP+TATA[1]. Of the screened kernels, the linear type successfully classified each of the interactions (test accuracy =100%) and was the simplest of the considered kernel types. Such accuracy was not surprising given the large change in R_{ct} for a protein-biomolecule interaction of this type (see **Fig 1**). Results associated with the other common, but more complex, kernels and their optimization parameters are shown in supplemental Figure S3-S4.

The molecular interactions between TBP and MBF/TATA[1] shown in **Fig 1** can be viewed as between transient and permanent interactions[52]. For these moderate to tight biomolecule interactions, a basic Randles-Ershler equivalent circuit or SVM (linear kernel) analysis can be used to analyze the data. In the following section, we show that for analysis of much weaker reversible protein-ligand interactions, equivalent circuit analysis is not sufficient and more complicated SVM classification must be used for accurate analysis.
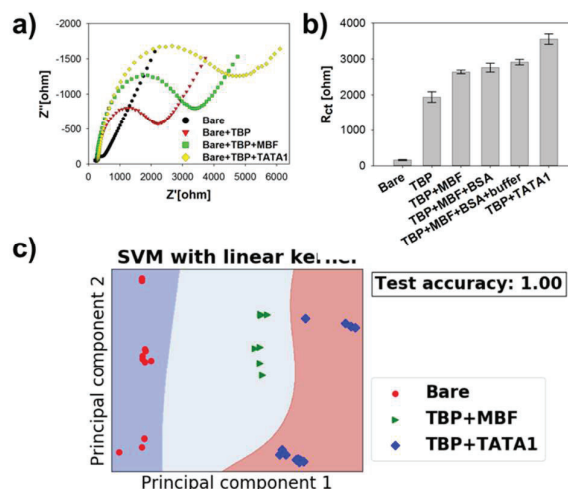

Figure 1. Impedimetric biosensor for the detection of protein-protein or protein-DNA interactions. **a)** Representative Nyquist plots for TBP-MBF interactions or TBP-TATA interactions clearly show an increase in charge transfer resistance after addition of target. **b)** Average R_{ct} derived from Randles-Ershler equivalent circuit model show significant results for TBP-biomolecule interactions but no significant change after addition of buffer or BSA (p values shown for each test group). **c)** Support vector machine (SVM) classification results with linear kernel. Additional Nyquist plots, average charge transfer resistance, and other SVM kernels are shown in the supplemental section.

Reversible protein-ligand interactions

To further challenge the machine learning tool, a CSP biosensor for detecting acetone at levels relevant for DKA triage diagnosis was developed and tested. DKA is a potentially fatal outcome from

complications associated with diabetes, and accurate measurement of acetone is challenging. CSP are an excellent candidate for binding volatiles such as acetone, but to date the technology has not been proven. Representative Nyquist plots (**Fig 2a**) and Bode plots (**Fig 2b**) show that the adsorption of CSP onto the electrode caused a significant change in EIS spectra, but the change after addition of clinically relevant acetone (5 mM) was less pronounced.

For a more detailed *post hoc* analysis, a Randles-Ershler equivalent circuit was used to derive R_s, R_{ct}, Z_w, and C_{dl} as previously described (**Fig 2c**). In addition, net impedance at various cut-off frequencies was extracted from Bode plots (**Fig 2d**). Using a 99.9% confidence level, there was no significant difference between baseline measurements and average R_s (p=0.015), R_{ct} (p=0.002) Z_w (p=0.016), or impedance at any cut-off frequency (p<0.002) after addition of 5mM acetone.
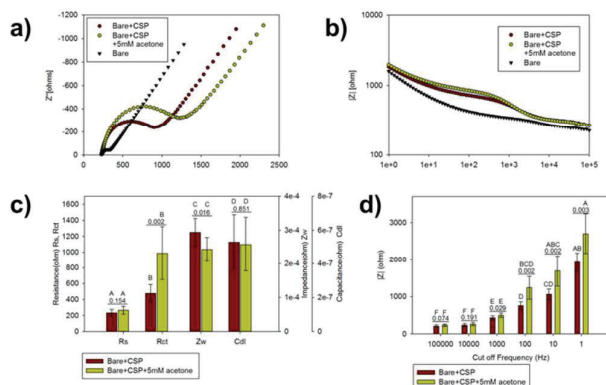

Figure 2. EIS analysis of CSP-acetone interactions in the presence and absence of 5mM acetone. Representative a) Nyquist plots and b) Bode plots. c) Average parameters from Randles-Ershler equivalent circuit analysis (R_s, R_{ct}, Z_w and C_{dl}) from Randel's equivalent circuit, baseline data and EIS in the presence of 5mM acetone. d) Net impedance at representative cut off frequencies. In panels ac and d, numbers denote the p-value and uppercase letters denote statistically significant groups.

To further analyze the spectra, more complex equivalent circuit models were analyzed using ZMAN software with Chi^2 fitting. All equivalent circuits with improved Chi^2 fit (relative to Randles-Ershler) had more than four elements in various parallel/series connections, including at least one resistive element(R), capacitive element(C), constant phase element (Q) and inductive element (I) (see supplemental Fig S5). However, statistical analysis of the output parameters for these circuits also showed no significant difference in baseline and in the presence of acetone for replicate biosensors. Furthermore, there is no direct physical analogous biological structure to the constant phase elements (Q) produced by the model, further complicating the interpretation of the results and inducing bias on the interpretation.

The case study in **Fig 2** represents a common issue in non-Faradaic impedimetric biosensing where the device is based on interaction of proteins and small molecules. In such a case, the individual biosensor responds to target analyte, but variability of replicate sensors is high and interpretation of results at relevant levels is challenging. This is particularly true for weak/reversible interactions between small molecules and proteins where there is not an inherent reaction (as is the case for CSP-ligand binding). The CSP biosensor system is a promising biomimetic sensor system, but

more accurate p*ost hoc* tools are needed for accurate detection of target biomarkers. As described by Liu et al [53], the underlying cause for this challenge is likely a result of the nature of CSP-ligand binding in sensors. Liu et al showed that protein conformation change (backbone displacement) plays a major role in the electrical (Faradaic) properties of the sensor; this work was based on the honeybee protein Ac-ASP3. Since conformation changes can occur with non-specific interactions such as hydrogen bonding, CSP biosensors are subject to erroneous outputs due to non-specific interactions. To alleviate the false negative issue shown in **Fig 2**, EIS data was further analyzed by SVM classification.

SVM classification for acetone-CSP interactions

The decision boundaries for each kernel are shown in **Fig 3**, where testing data that fall into blue areas is predicted as negative (no acetone) and those that fall in red areas as positive ($\geq$ 5mM acetone). As discussed by Liu et al [54], other *post hoc* algorithms not analyzed here, such as random forest, may be more accurate in some cases. However, these approaches often increase accuracy by overfitting the data[55], which ultimately decreases the robustness of the classifier. Moreover, many of these are computationally expensive and cannot be analyzed using mobile hardware such as a mobile phone or tablet. The Gaussian radial base function (RBF) kernel (accuracy = 98%) had the highest test accuracy for classifying the training dataset. However, using the default kernel settings the dataset was not linearly separable and the RBF kernel had an overfitting issue, requiring further analysis and tuning of the parameters.
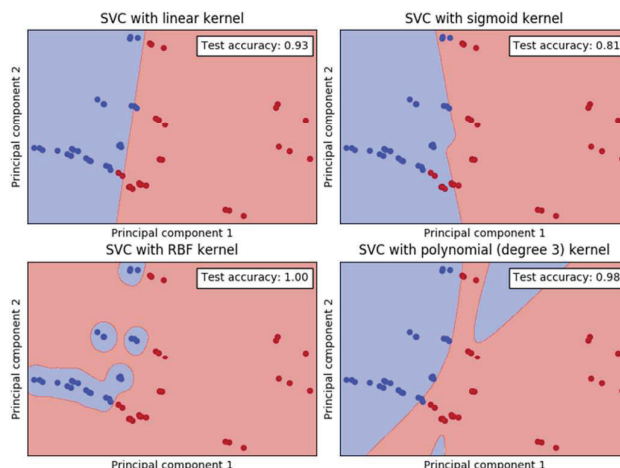

Figure 3. SVM classification for CSP-acetone biosensors using four common kernels. **a)** linear kernel (test accuracy =96%), **b)** sigmoidal kernel (test accuracy =83%), **c)** radial base function kernel (test accuracy =98%), and **d)** polynomial kernel (test accuracy =96%). Blue dots represented baseline EIS signals (no acetone in samples) and red dots represented positive EIS signals (5mM acetone in samples). The decision surface of these four SVM classifiers are plotted by red and blue regions.

To tune the RBF kernel parameters, a grid search and cross validation were performed. In cross validation, the original dataset was shuffled and divided into ten different training and testing sets, with 20% of the total data used for testing. Next, each training set was used to fit the SVM classifier and average test accuracy calculated for each split training set. In the RBF kernel, the two governing hyper-parameters are the

penalty parameter (C) and non-linear kernel coefficient (γ). The penalty hyper-parameter trades off misclassification against simplicity of decision surface, where lower C values tolerate more mistakes. The non-linear parameter defines the influence of a single training example on the output, and can be seen as the inverse of the radius of the influence of support vectors[50]. Each of these parameters were optimized using a grid parameter search function using the RBF kernel (**Fig 4**). In the top left panel of **Fig 4**, where C is low, the penalty for misclassification is small and the decision surface is simple relative to values in the first column with higher C values. As the nonlinear hyper-parameter increases (from left to right in **Fig 4**), the influence radius decreases, causing over-fitting. The protocol described herein resolves this issue by creating a visualization tool to select the optimum hyper-parameters.
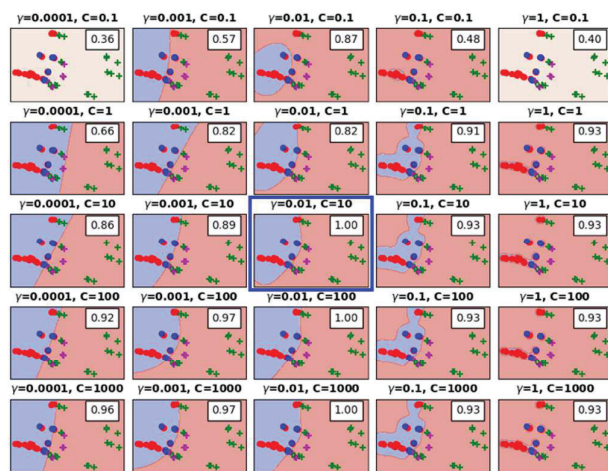

Figure 4. Tuning of RBF hyper-parameters (C and gamma) for CSP acetone interactions. Representative SVM classification results for one training and testing set show the effects of parameters C and g in the output of the RBF kernels. Red and blue circles represent the baseline samples in training and testing sets; green and purple plus symbols represent the positive signals in training and testing sets. The background blue and red region indicated the classifier decision surface, where all data fall into the red region are predicted as positive. Cross-validation scores are shown in the top right corner of each subplot. The optimal classifier zone is highlighted with a blue rectangle in the center of the image.

The Python code has a built-in function to optimize the hyper-parameters from data such as that shown in **Fig 4**. Based on this heat map (**Fig 5**), the optimum value of γ was 0.01, and the optimum value of C was 10. Using these parameters, the Python code is then modified (see details in step-by-step user guide) and the data is analyzed. Using the optimized kernel selection and hyper-parameters, the SVM demonstrated an accuracy of 95 ± 4% in cross validation and prediction of test samples.
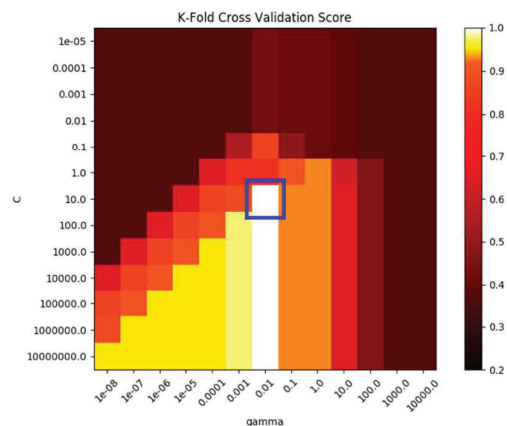

Figure 5. Heatmap of validation accuracy as a function of RBF parameters C and γ. The color indicates the cross-validation accuracy, where lighter colors represent a higher cross-validation score. Optimal parameters are highlight with a blue rectangle in the center

Decomposition of high dimensional EIS data to two-dimensional data with PCA is known to improve sensor/detector accuracy due to identification of uncorrelated variables from a large set of data[56]. PCA explains the maximum amount of data variance with the fewest number of principal components. For a semi-quantitative biosensor application such as the data in **Fig 1-5**, the use of only two principal components can lead to loss of useful information during data decomposition. However, for the RFP kernel with optimal tuning parameters the results were statistically significant at the 95% confidence interval. To further analyze the dataset, classifiers with 3 and 10 principal components were built and the cross-validation accuracy was improved (97 ± 3%), which is expected as less information was lost during decomposition (a 3D data representation for data analyzed with three principal components is shown in supplemental Fig S6). This result was expected, as use of reductionist clustering (i.e., using two-dimensional PCA) increases the risk of eliminating important outliers within the data. For example, over-clustering could result in important deviations from the "normal", for example in the case of silent ischemia[57]. In this case the data curation can be improved by analyzing polar coordinates in lieu of, or in addition to, Cartesian coordinates from impedimetric sensor data. However, analysis of classifiers with a dimension larger than two is computationally expensive, and can make use of mobile phone based analytical systems challenging. Care should be taken to discern as to whether the computational need outweighs the ability to analyze data on site using mobile equipment such as a tablet or mobile phone. To maintain focus on mobile-enabled diagnostic systems in this study, we used a two-dimensional PCA analysis, which is valid for semi-quantitative biosensor data where a regulatory or diagnostic metric is known (such as the case of DKA salivary biomarkers shown here).

Although not used here, computational speed and memory requirement can be improved by using more advanced computational tools such as the tensor compiler by Kjolstad et al[58]. This approach is particularly useful for multidimensional data analysis, and provides a generic mechanism that can

generate code for compound tensor operations with sparse tensors, eliminating the need for writing optimized code for a specific problem. This tensor algebra compiler library represents an excellent next step forward to improve the work herein.

The SVM tool shown here is highly useful for point of need small molecule analysis using mobile detection and analysis systems (see supplemental Figure S7). Rapid triage analysis of breath biomarkers (including acetone and β-hydroxybutyrate) is vital for triage analysis of patients with DKA symptoms, and mobile phone solutions can bring this diagnosis to rural areas where health care is limited. The overall mortality rate for DKA ranges from 1 to 10% of all patient admissions, and an even higher mortality rate is found among non-hospitalized patients and children under the age of 10 [59]. Convergent technologies for triage diagnostics require systems-level solutions that are based on readily accessible hardware such as mobile phones or tablets [60].

Conclusions

Biosensors based on weak/transient interactions between small molecules and bioreceptors are a challenge for detection electronics, particularly in field studies or in analysis of complex matrices (e.g., body fluids, food, river water, etc.) using non-Faradaic impedimetric sensors. Support vector machine learning tools are facile *post hoc* analysis tools that do not require significant computational power and can be used for *in situ* analysis with mobile hardware such as a mobile phone or tablet. Here, we show use of a simple, open source machine learning algorithm for analysing such impedimetric data, and we show that the tool can be used for point of need applications.

Conflicts of interest

There are no conflicts to declare.
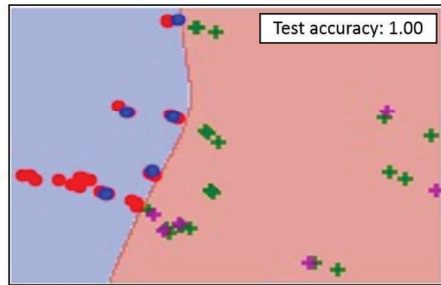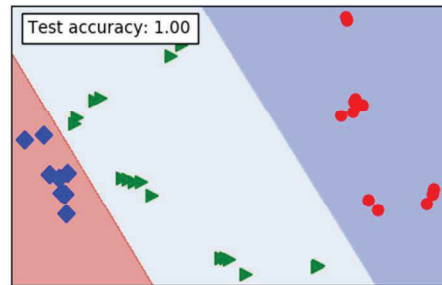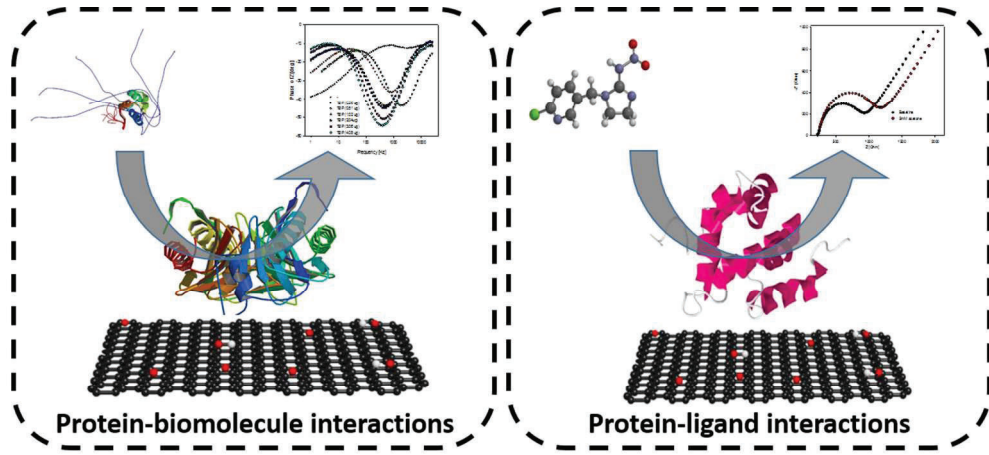
Acknowledgements

References

1. A. P. F. Turner, *Chemical Society reviews*, 2013, **42**, 3184-3196.
2. A. R. Demaio and J. Rockstroem, *Lancet*, 2015, **386**, E36-E37.
3. D. Z. Ji, L. Liu, S. Li, C. Chen, Y. L. Lu, J. J. Wu and Q. J. Liu, *Biosens Bioelectron*, 2017, **98**, 449-456.
4. L. Liu, D. M. Zhang, Q. Zhang, X. Chen, G. Xu, Y. L. Lu and Q. J. Liu, *Biosens Bioelectron*, 2017, **93**, 94-101.
5. D. M. Zhang, J. Jiang, J. Y. Chen, Q. Zhang, Y. L. Lu, Y. Yao, S. Li, G. L. Liu and Q. J. Liu, *Biosens Bioelectron*, 2015, **70**, 81-88.
6. A. Hayat and J. L. Marty, *Front Chem*, 2014, **2**.
7. D. Vanegas, C. Gomes and E. McLamore, *Biosens J*, 2016, **5**, 2.
8. A. Walcarius, S. D. Minteer, J. Wang, Y. Lin and A. Merkoçi, *J Mater Chem B*, 2013, **1**, 4878-4908.
9. A. Bonanni, A. H. Loo and M. Pumera, *TrAC Trends in Analytical Chemistry*, 2012, **37**, 12-21.
10. T. Yin and W. Qin, *TrAC Trends in Analytical Chemistry*, 2013, **51**, 79-86.
11. M. A. Daniele, M. Pedrero, S. Burrs, P. Chaturvedi, W. W. A. Wan Salim, F. Kuralay, S. Campuzano, E. McLamore, A. A. Cargill, S. Ding and J. C. Claussen, in *Nanobiosensors and Nanobioanalyses*, eds. M. d. C. Vestergaard, K. Kerman, I. M. Hsing and E. Tamiya, Springer Japan, Tokyo, 2015, DOI: 10.1007/978-4-431-55190-4_8, pp. 137-166.
12. C. Z. Zhu, G. H. Yang, H. Li, D. Du and Y. H. Lin, *Anal Chem*, 2015, **87**, 230-249.
13. E. S. McLamore, M. Convertino, I. Ocsoy, D. C. Vanegas, M. Taguchi, Y. Rong, C. Gomes, P. Chaturvedi and J. C. Claussen, in *Semiconductor-Based Sensors*, WORLD SCIENTIFIC, 2016, DOI: 10.1142/9789813146730_0002, pp. 35-67.
14. J. S. Daniels and N. Pourmand, *Electroanal*, 2007, **19**, 1239-1257.
15. E. B. Bahadir and M. K. Sezginturk, *Artif Cells Nanomed Biotechnol*, 2016, **44**, 248-262.
16. M. I. Prodromidis, *Electrochim Acta*, 2010, **55**, 4227-4233.
17. J.-G. Guan, Y.-Q. Miao and J.-R. Chen, *Biosensors and Bioelectronics*, 2004, **19**, 789-794.
18. R. Elshafey, A. C. Tavares, M. Siaj and M. Zourob, *Biosens Bioelectron*, 2013, **50**, 143-149.
19. D. D. Macdonald, *Electrochim Acta*, 2006, **51**, 1376-1388.
20. M. E. Orazem, P. Agarwal and L. H. Garciarubio, *J Electroanal Chem*, 1994, **378**, 51-62.
21. S. R. Das, Q. Nian, A. A. Cargill, J. A. Hondred, S. W. Ding, M. Saei, G. J. Cheng and J. C. Claussen, *Nanoscale*, 2016, **8**, 15870-15879.
22. W. J. Yuan, Y. Zhou, Y. R. Li, C. Li, H. L. Peng, J. Zhang, Z. F. Liu, L. M. Dai and G. Q. Shi, *Sci Rep-Uk*, 2013, **3**.
23. I. I. Suni, *TrAC Trends in Analytical Chemistry*, 2008, **27**, 604-611.
24. H. Song, Y. Wang, J. M. Rosano, B. Prabhakarpandian, C. Garson, K. Pant and E. Lai, *Lab Chip*, 2013, **13**, 2300-2310.
25. R. Kumar, A. P. Bhondekar, R. Kaur, S. Vig, A. Sharma and P. Kapur, *Sensor Actuat B-Chem*, 2012, **171**, 1046-1053.
26. Q. Dong, L. Du, L. Zhuang, R. Li, Q. Liu and P. Wang, *Biosens Bioelectron*, 2013, **49**, 263-269.
27. L. Lu, S. P. Deng, Z. W. Zhu and S. Y. Tian, *Food Analytical Methods*, 2015, **8**, 1893-1902.
28. Y. X. Dai, X. Wang, P. B. Zhang and W. H. Zhang, *Measurement*, 2017, **109**, 408-424.
29. X. Ding, Z. Lv, C. Zhang, X. Gao and B. Zhou, *IEEE Access*, 2017.
30. Z. Qin, B. Zhang, K. Gao, L. Zhuang, N. Hu and P. Wang, *Sensors and Actuators B: Chemical*, 2017, **239**, 746-753.
31. O.-P. Smolander, A. S. Ribeiro, O. Yli-Harja and M. Karp, *Sensors and Actuators B: Chemical*, 2009, **141**, 604-609.
32. E. Akbari, Z. Buntat, E. Shahraki, R. Parvaz and M. J. Kiani, *Journal of biomaterials applications*, 2016, **30**, 677-685.

33. F. F. Gonzalez-Navarro, M. Stilianova-Stoytcheva, L. Renteria-Gutierrez, L. A. Belanche-Muñoz, B. L. Flores-Rios and J. E. Ibarra-Esquer, *Sensors-Basel*, 2016, **16**, 1483.

34. Z. Q. Geng, S. S. Zhao, G. C. Tao and Y. M. Han, *Food Control*, 2017, **78**, 33-42.

35. S. De Vito, E. Esposito, M. Salvato, O. Popoola, F. Formisano, R. Jones and G. Di Francia, *Sensors and Actuators B: Chemical*, 2017, DOI: https://doi.org/10.1016/j.snb.2017.07.155.

36. O. Sadik, W. H. Land, A. K. Wanekaya, M. Uematsu, M. J. Embrechts, L. Wong, D. Leibensperger and A. Volykin, *Journal of chemical information and computer sciences*, 2004, **44**, 499-507.

37. T. Alizadeh and S. Zeynali, *Sensors and Actuators B: Chemical*, 2008, **129**, 412-423.

38. Y. Zuo, S. Chakrabartty, Z. Muhammad-Tahir, S. Pal and E. C. Alocilja, *Ieee Sens J*, 2006, **6**, 1644-1651.

39. R. G. Vogt, in *Comprehensive Molecular Insect Science*, Elsevier, Amsterdam, 2005, DOI: https://doi.org/10.1016/B0-44-451924-6/00047-8, pp. 753-803.

40. A. Sanchez-Gracia, F. G. Vieira and J. Rozas, *Heredity*, 2009, **103**, 208-216.

41. D. C. Vanegas, M. Taguchi, P. Chaturvedi, S. Burrs, M. Tan, H. Yamaguchi and E. S. McLamore, *Analyst*, 2014, **139**, 660-667.

42. S. L. Burrs, D. C. Vanegas, Y. Rong, M. Bhargava, N. Mechulan, P. Hendershot, H. Yamaguchi, C. Gomes and E. S. McLamore, *Analyst*, 2015, **140**, 1466-1476.

43. C. Song, A. Ortiz-Urquiza, S. H. Ying, J. X. Zhang and N. O. Keyhani, *Plos One*, 2015, **10**.

44. Y. Rong, J. Kieran-Lewis, N. O. Keyhani and E. S. McLamore, 2016.

45. S. Fujii, T. Maeda, I. Noge, Y. Kitagawa, K. Todoroki, K. Inoue, J. Z. Min and T. Toyo'oka, *Clin Chim Acta*, 2014, **430**, 140-144.

46. C.-W. Hsu, C.-C. Chang and C.-J. Lin, 2003.

47. T. Hastie, R. Tibshirani and J. Friedman, in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer New York, New York, NY, 2009, DOI: 10.1007/978-0-387-84858-7_7, pp. 219-259.

48. N. Halko, P. G. Martinsson and J. A. Tropp, *Siam Rev*, 2011, **53**, 217-288.

49. J. Bergstra and Y. Bengio, *J Mach Learn Res*, 2012, **13**, 281-305.

50. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J Mach Learn Res*, 2011, **12**, 2825-2830.

51. S. W. Ding, C. Mosher, X. Y. Lee, S. R. Das, A. A. Cargill, X. H. Tang, B. L. Chen, E. S. McLamore, C. Gomes, J. M. Hostetter and J. C. Claussen, *Acs Sensors*, 2017, **2**, 210-217.

52. I. M. Nooren and J. M. Thornton, *Journal of molecular biology*, 2003, **325**, 991-1018.

53. Q. J. Liu, H. Wang, H. L. Li, J. Zhang, S. L. Zhuang, F. N. Zhang, K. J. Hsia and P. Wang, *Biosens Bioelectron*, 2013, **40**, 174-179.

54. M. Liu, M. Wang, J. Wang and D. Li, *Sensors and Actuators B: Chemical*, 2013, **177**, 970-980.

55. T. Hastie, R. Tibshirani and J. Friedman, in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer New York, New York, NY, 2009, DOI: 10.1007/978-0-387-84858-7_2, pp. 9-41.

56. H. Abdi and L. J. Williams, *Wiley interdisciplinary reviews: computational statistics*, 2010, **2**, 433-459.

57. P. F. Cohn, K. M. Fox and C. Daly, *Circulation*, 2003, **108**, 1263-1277.

58. F. Kjolstad, S. Kamil, S. Chou, D. Lugato and S. Amarasinghe, 2017.

59. R. Ganesh, N. Suresh and J. Ramesh, *The National medical journal of India*, 2006, **19**, 155-158.

60. R. Chiu, C. Ho, S. Tong, K. Ng and C. Lam, *Hong Kong medical journal= Xianggang yi xue za zhi/Hong Kong Academy of Medicine*, 2002, **8**, 172-176.

Figure 1. Impedimetric biosensor for the detection of protein-protein or protein-DNA interactions. a) Representative Nyquist plots for TBP-MBF interactions or TBP-TATA interactions clearly show an increase in charge transfer resistance after addition of target. b) Average Rct derived from Randles-Ershler equivalent circuit model show significant results for TBP-biomolecule interactions but no significant change after addition of buffer or BSA (p values shown for each test group). c) Support vector machine (SVM) classification results with linear kernel. Additional Nyquist plots, average charge transfer resistance, and other SVM kernels are shown in the supplemental section

Figure 2. EIS analysis of CSP-acetone interactions in the presence and absence of 5mM acetone. Representative a) Nyquist plots and b) Bode plots. c) Average parameters from Randles-Ershler equivalent circuit analysis (Rs, Rct, Zw and Cdl) from Randel's equivalent circuit, baseline data and EIS in the presence of 5mM acetone. d) Net impedance at representative cut off frequencies. In panels ac and d, numbers denote the p-value and uppercase letters denote statistically significant groups.

Figure 3. SVM classification for CSP-acetone biosensors using four common kernels. a) linear kernel (test accuracy =96%), b) sigmoidal kernel (test accuracy =83%), c) radial base function kernel (test accuracy =98%), and d) polynomial kernel (test accuracy =96%). Blue dots represented baseline EIS signals (no acetone in samples) and red dots represented positive EIS signals (5mM acetone in samples). The decision surface of these four SVM classifiers are plotted by red and blue regions.

203x151mm (100 x 100 DPI)

Figure 4. Tuning of RBF hyper-parameters (C and gamma) for CSP acetone interactions. Representative SVM classification results for one training and testing set show the effects of parameters C and g in the output of the RBF kernels. Red and blue circles represent the baseline samples in training and testing sets; green and purple plus symbols represent the positive signals in training and testing sets. The background blue and red region indicated the classifier decision surface, where all data fall into the red region are predicted as positive. Cross-validation scores are shown in the top right corner of each subplot. The optimal classifier zone is highlighted with a blue rectangle in the center of the image.

203x151mm (100 x 100 DPI)
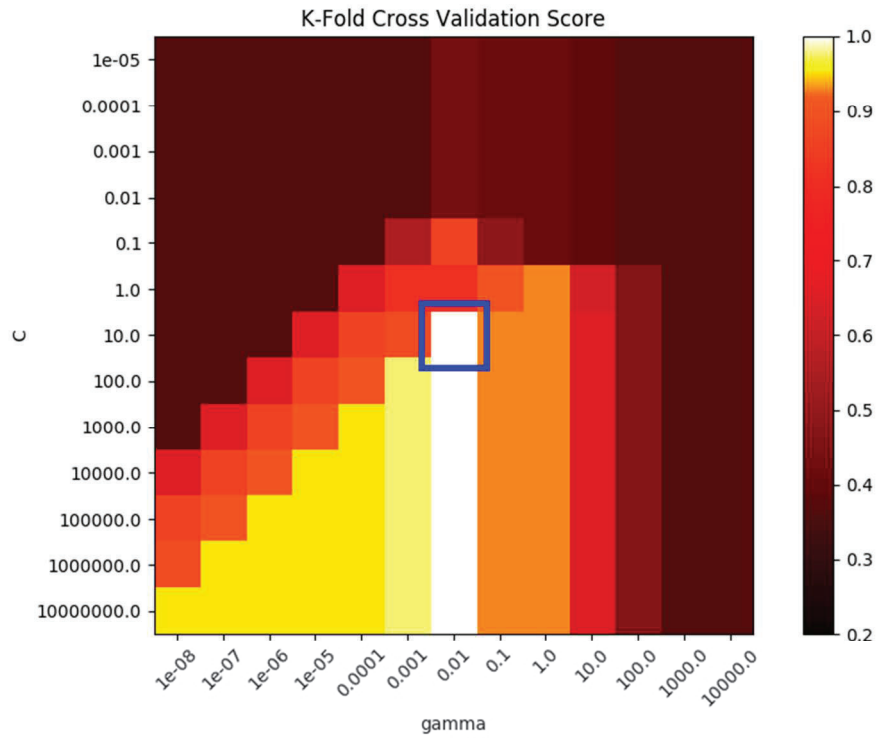
Figure 5. Heatmap of validation accuracy as a function of RBF parameters C and γ. The color indicates the cross-validation accuracy, where lighter colors represent a higher cross-validation score. Optimal parameters are highlight with a blue rectangle in the center

203x151mm (100 x 100 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SUPPLEMENTAL

Table S1. Randles-Ershler equivalent circuit analysis results for TBP-MBF binding and TBP-TATA[1] binding.

Binding	Circuit element	Value (ohm)	Standard deviation
TBP	R_s	265.73	0.38
	R_{ct}	2021.65	45.63
	Z_w	1.80E-04	6.38E-07
	C_{dl}	5.10E-07	4.58E-10
TBP-MBF	R_s	266.50	1.78
	R_{ct}	2640.88	52.30
	Z_w	1.80E-04	1.85E-06
	C_{dl}	4.91E-07	8.03E-09
TBP-TATA1	R_s	328.45	2.78
	R_{ct}	3395.19	302.53
	Z_w	1.41E-04	1.46E-05
	C_{dl}	5.85E-07	2.75E-08

Figure S1. Piture of a 3D printed plastic cap fitted onto standard Basi Pt/Ir electrode, and engineering drawing design of the cap.
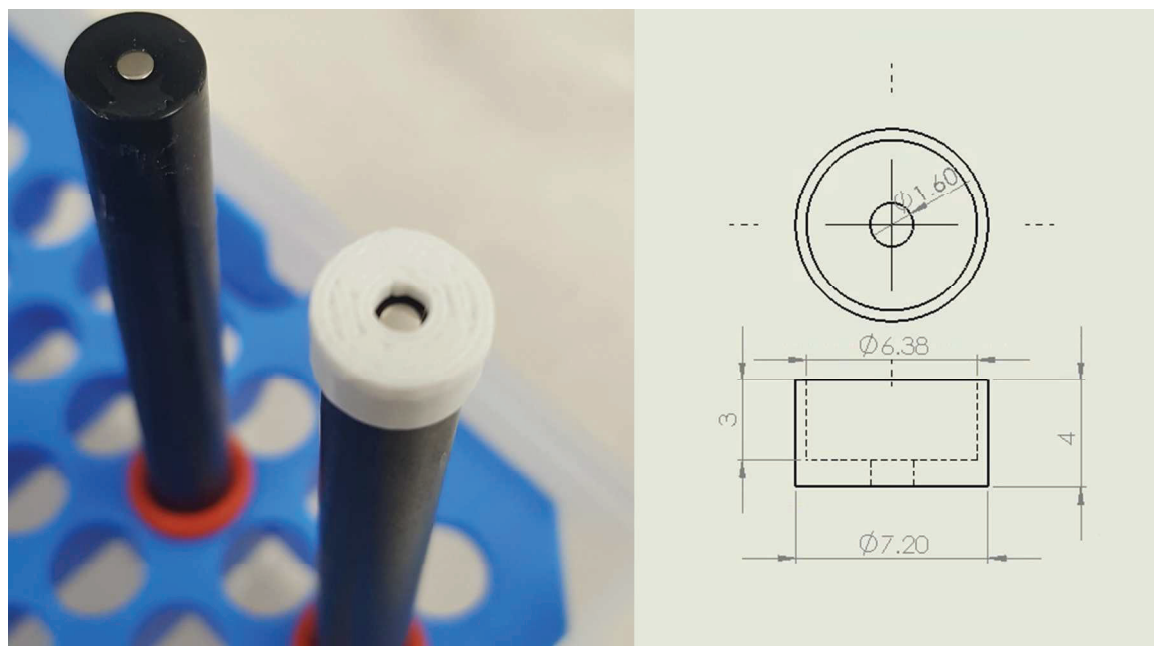
Figure S2. Charger transfer resisitance (R_{ct}) from a Randel's model for addition of BSA or buffer to TBP coupled electrodes and TBP-MBF coupled electrodes.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
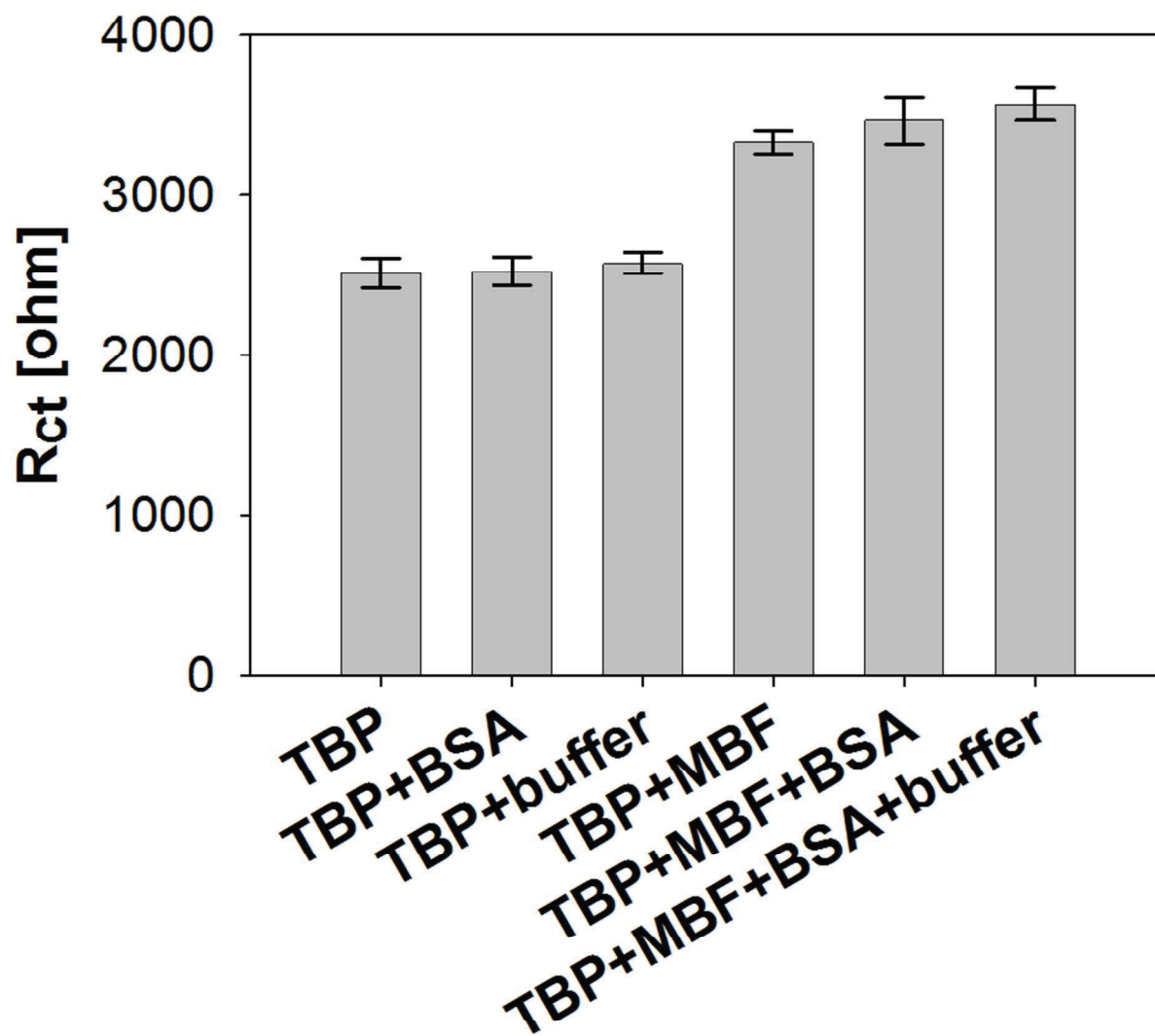
Figure S3. SVM results for TBP, TBP+MBF and TBP+TATA[1] calasification with common kernels.
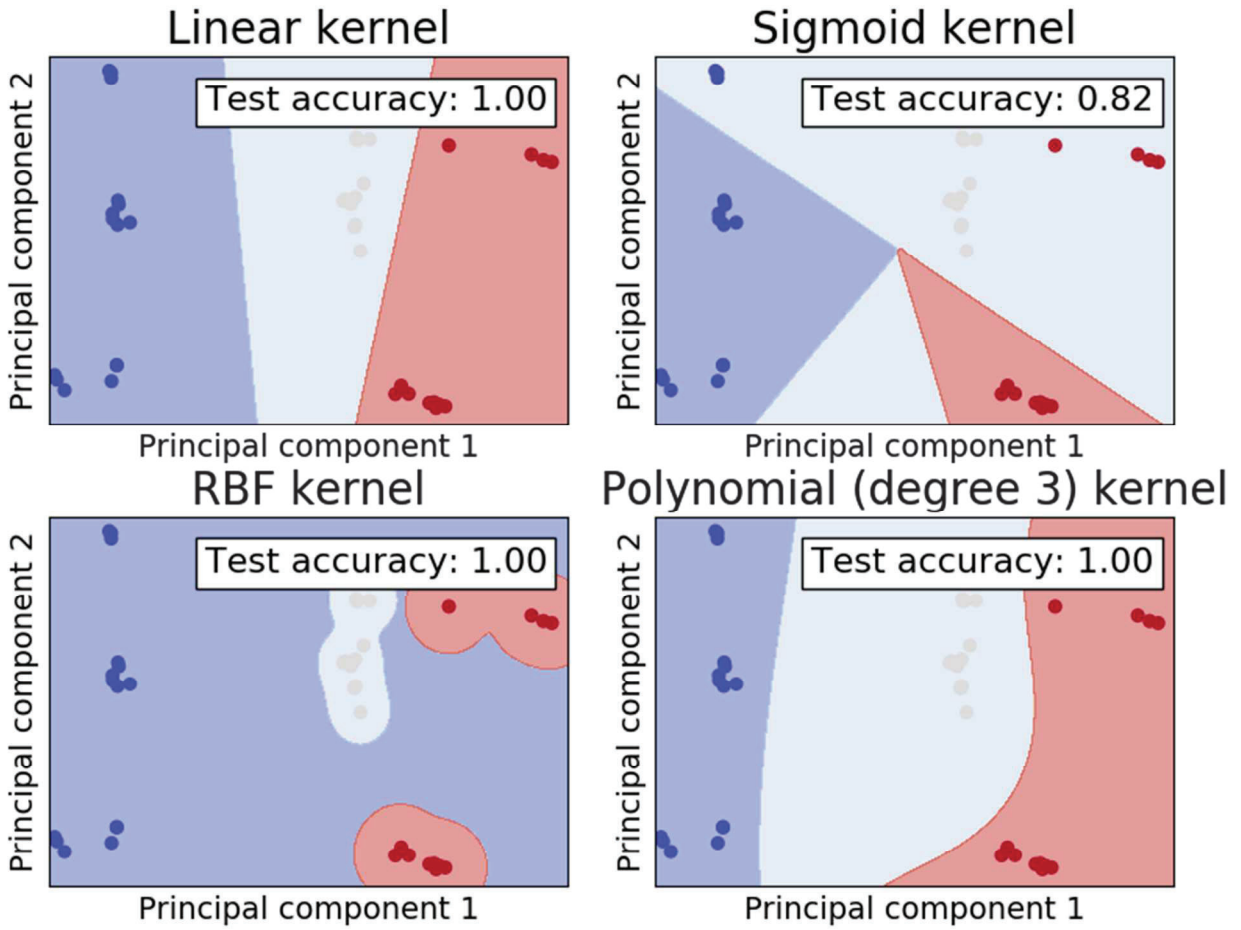
Figure S4. Representative SVM parameter tunning result for TBP, TBP+MBF and TBP+TATA[1] calasification with RBF kernel.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure S5. Representative equivalent circuit models tested with model search function in ZMAN software.
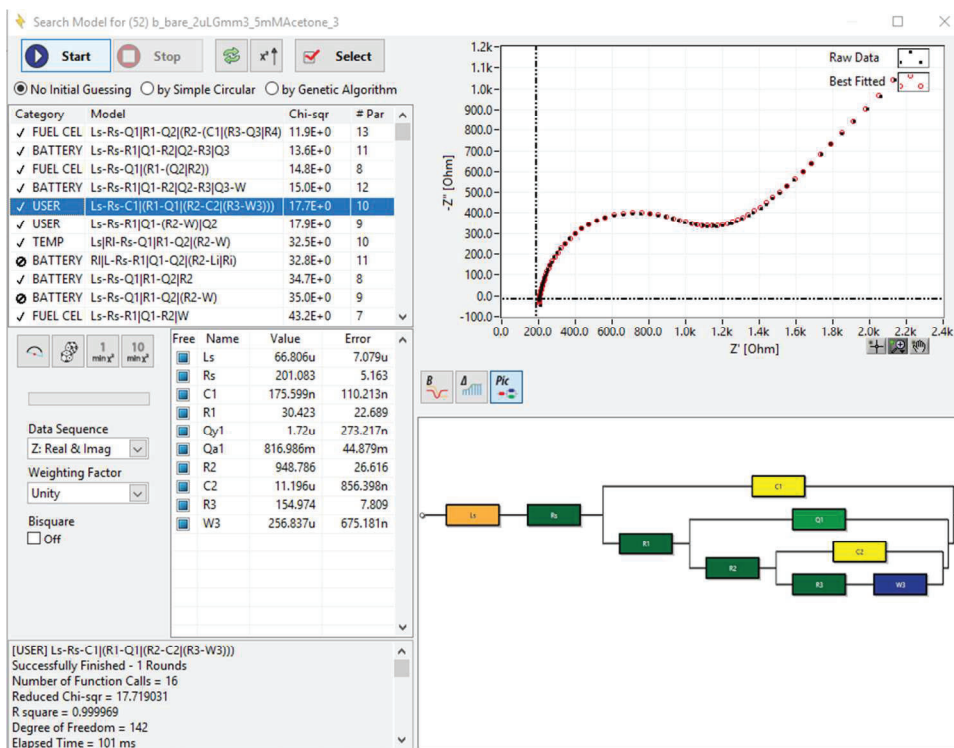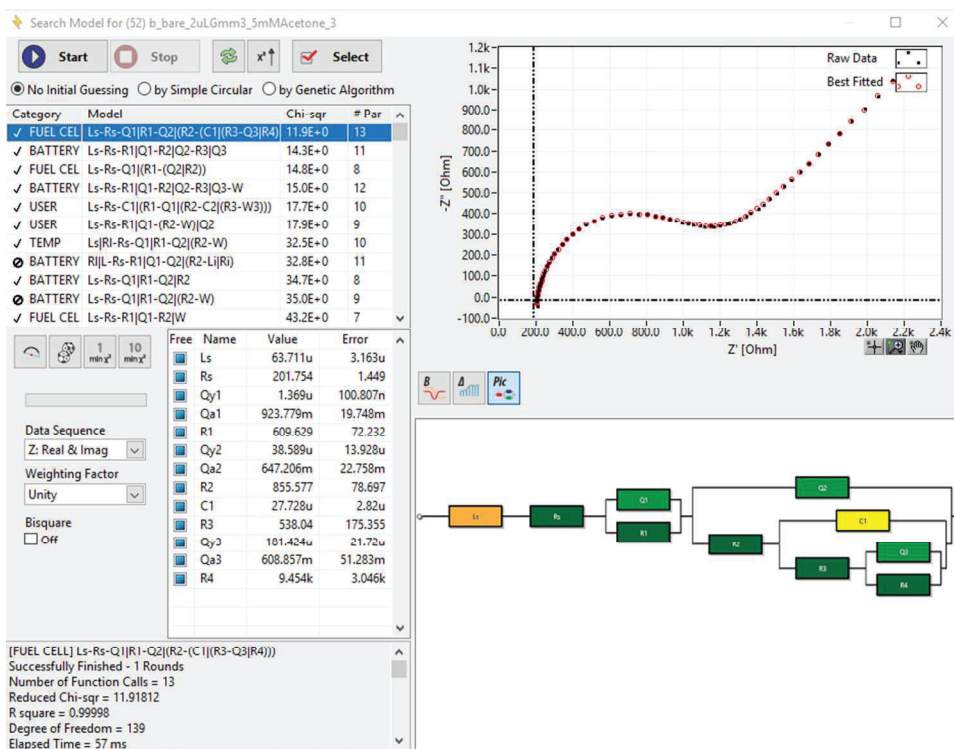
Figure S6. 3D data representation for data analyzed with three principal components. Red dots represented baseline signals of Gmm CSP biosensor, green plus represented 5mM acetone detection signals.
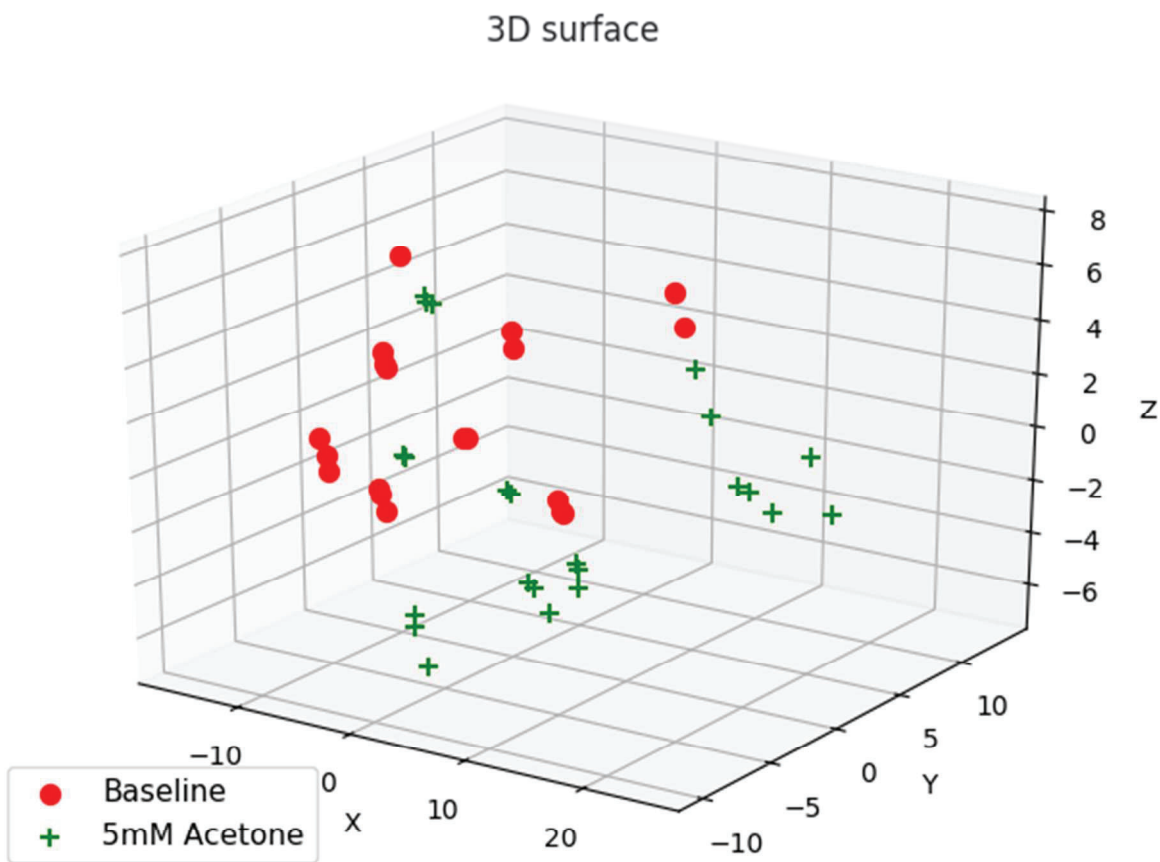
Figure S7. Support vector machine learning software installed on Android phone with small molecule using mobile detection and analysis systems. Screenshot showed Jupyter notebook running a SVM kernel selection code.
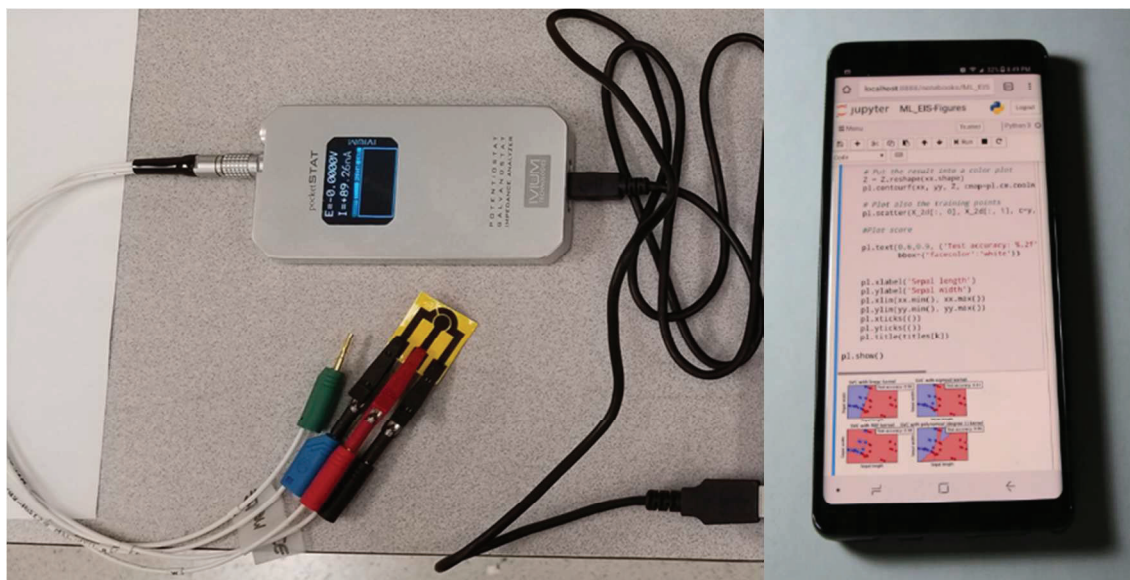
SVM Machine Learning for EIS data analysis step-by-step user guide

Yue Rong, Agricultural and Biological Engineering, University of Florida

1. Software preparation

1.1. Download Anaconda installer

"Anaconda is a freemium open source distribution of the Python and R programming languages for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment." --https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)

It includes all the packages and tools we need and is completely free.

It can be downloaded at:

https://www.continuum.io/downloads

There are different versions of Anaconda for Windows, macOS and Linux; Python 3.6 and 2.7. It will be about 2 Gb after installation and downloading all packages. We suggest Python 3.6.
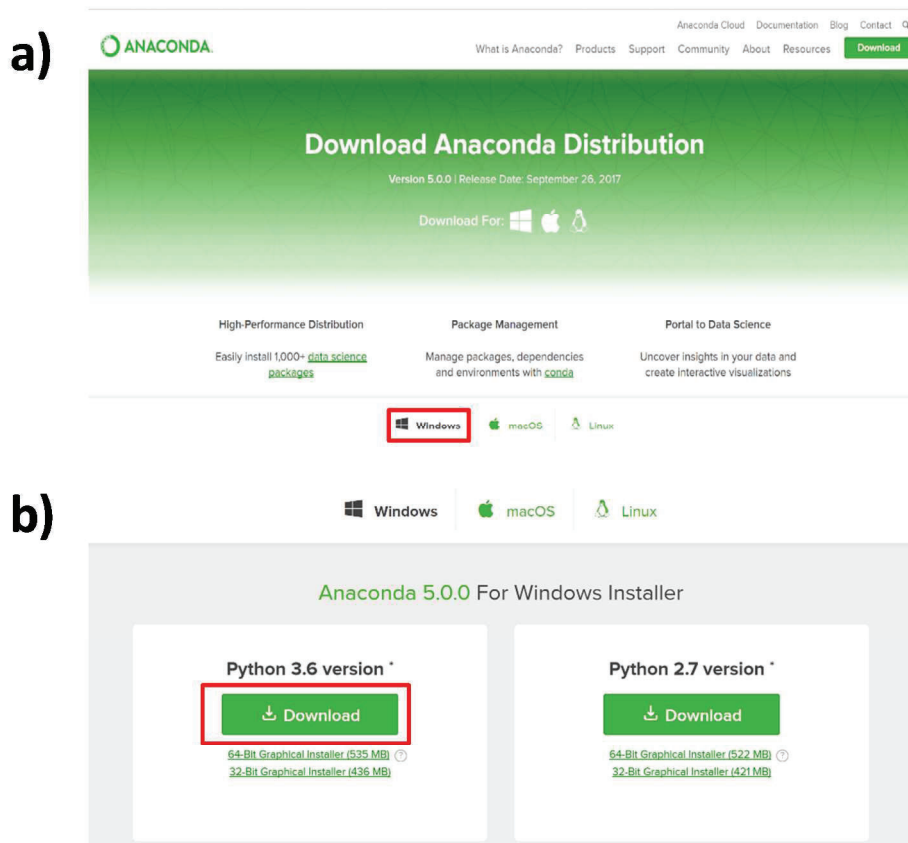
Figure U1. Download Anaconda

1.2. Install Anaconda

Double click the installer to launch

NOTE: If you encounter any issues during installation, temporarily disable your anti-virus software during install, then re-enable it after the installation concludes. If you have installed for all users, uninstall Anaconda and re-install it for your user only and try again.

Click Next.

Read the licensing terms and click "I Agree"

Select an install for "Just Me" unless you're installing for all users (which requires Windows Administrator privileges) and click Next.

Select a destination folder to install Anaconda and click the Next button.

NOTE: Install Anaconda to a directory path that does not contain spaces or unicode characters.

NOTE: Do not install as Administrator unless admin privileges are required.

Choose whether to add Anaconda to your PATH environment variable.

We recommend not adding Anaconda to the PATH environment variable, since this can interfere with other software. Instead, use Anaconda software by opening Anaconda Navigator or the Anaconda Prompt from the Start Menu.

Choose whether to register Anaconda as your default Python 3.6.

Unless you plan on installing and running multiple versions of Anaconda, or multiple versions of Python, you should accept the default and leave this box checked.

Click the Install button.

You can click Show Details if you want to see all the packages Anaconda is installing.

Click the Next button

After a successful installation, you will see the "Thanks for installing Anaconda" dialog box:

Click the Finish button

After your install is complete, verify it by opening Anaconda Navigator, a program that is included with Anaconda: from your Windows Start menu, select the shortcut Anaconda Navigator. If Navigator opens, you have successfully installed Anaconda.

Anaconda Navigator

If you had any issue installing the Anaconda, please visit https://docs.anaconda.com/anaconda/install/ for help.

2. Getting started

2.1. Open Anaconda Navigator

Find Anaconda 3 in "Windows Start Menu" and click "Anaconda Navigator" to open

2.2. Click "Launch" to open "Jupyter Notebook"

Figure 1 Lunch "Jupyter Notebook" in Anaconda Navigator

It will open in your internet browser (for me it's a Chrome) an the default file location is under your system user folder. (for example C:\Users\Username)
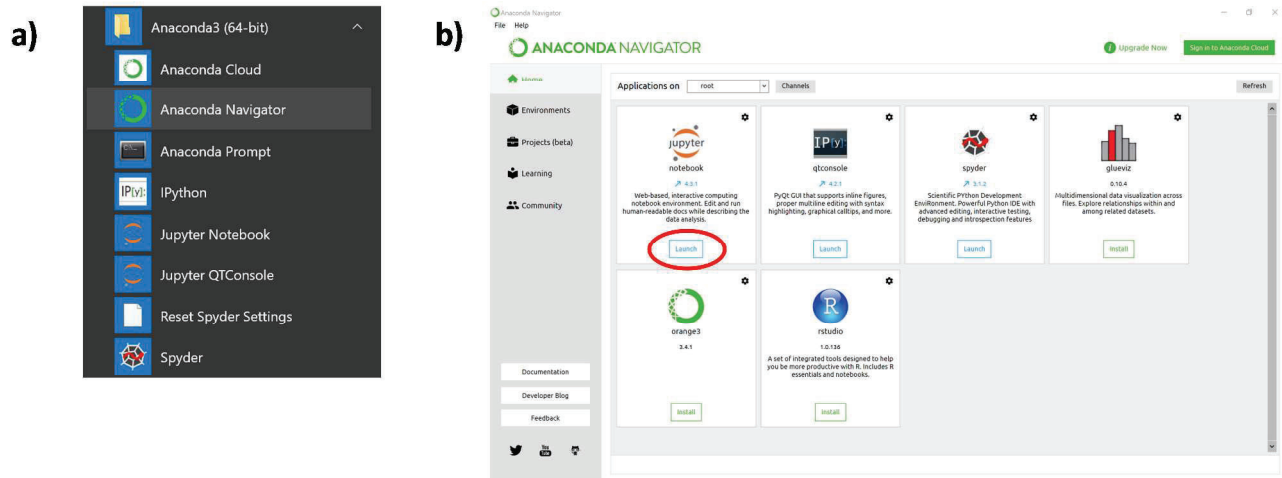
Figure U2. Open Jupyter Notebook

2.3. Run an example code

Download an example code from the link below and save it to "downloads" folder.

http://scikit-learn.org/stable/_downloads/plot_svm_nonlinear.ipynb

use the file browser to navigate and open the download source code above. It should look like Figure U3 below.
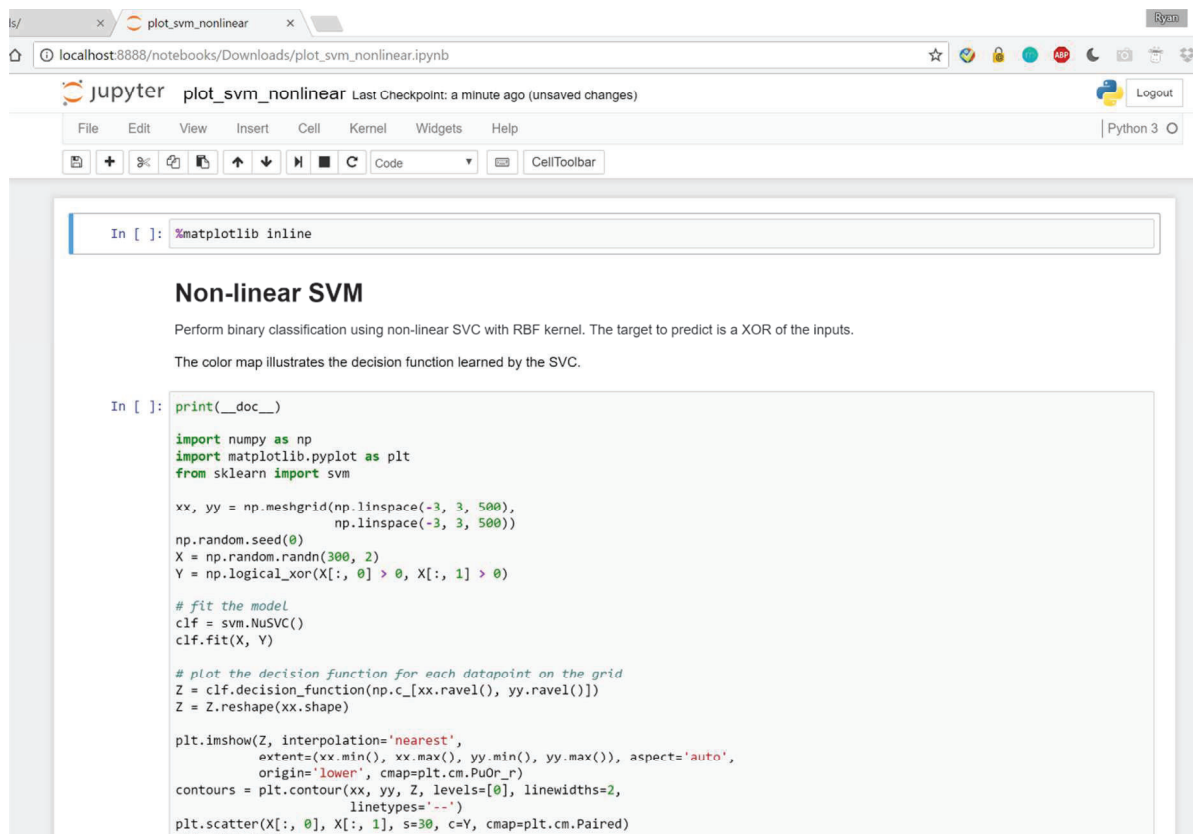
Figure U3. Example code for non-linear SVM

In "Cell" click "Run All" or run cells by clicking the [▶|] as shown in Figure U4.
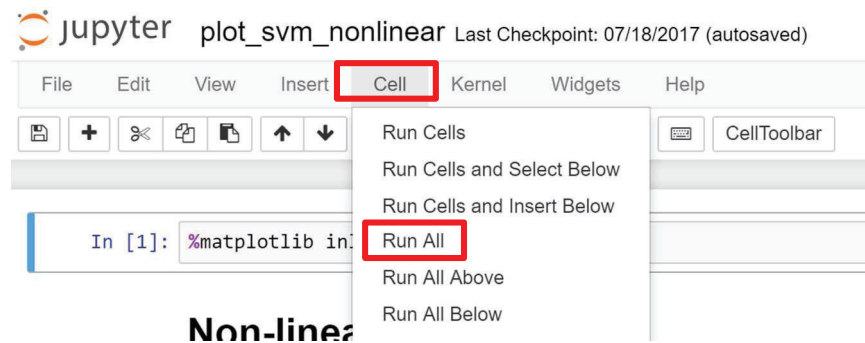
Figure U4. Run All cells to test the example code

The program should generate a plot similar to Figure U5.

```
plt.scatter(X[:, 0], X[:, 1], s=30, c=Y, cmap=plt.cm.Paired)
plt.xticks(())
plt.yticks(())
plt.axis([-3, 3, -3, 3])
plt.show()
```

Automatically created module for IPython interactive environment
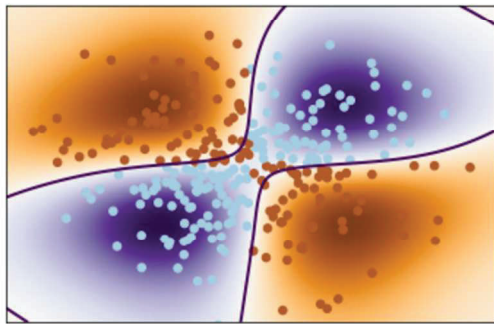
Figure U5. Expected output from example code

3. EIS data analysis with SVM machine learning

3.1. Data preparation

Raw EIS data usually includes frequency with real impedance (Z') and imaginary impedance (Z'') at that frequency as shown in Figure U6.

	A	B	C
1	Frequency [Hz]	Z'_raw [Ohm]	Z"_raw [Ohm]
0	99999.99984	258.054229	-7.637512
1	85769.58917	255.299691	-17.882566
2	73564.22411	278.492955	-18.88002
3	63095.73291	275.258838	-15.963164
4	54116.95083	289.347566	-28.062849
5	46415.88639	287.247927	-26.796648
6	39810.71513	286.058176	-32.585536
7	34145.48673	296.438676	-33.791812
8	29286.44376	301.039143	-34.728174
9	25118.8625	300.781714	-35.632715
10	21544.34515	307.305851	-40.781186
11	18478.4963	308.219676	-39.220288
12	15848.93039	312.550444	-44.296644
13	13593.56249	311.666707	-50.051774
14	11659.14268	315.616756	-54.547519
15	9999.998792	318.54106	-58.196081
16	8576.957885	322.457927	-63.004504
17	7356.421543	326.558636	-69.321829
18	6309.57253	330.949427	-77.043239
19	5411.69444	336.52842	-84.527607
20	4641.588089	343.795531	-92.286639
21	3981.071029	352.501566	-102.064143
22	3414.548269	359.595675	-112.278786
23	2928.644022	372.104547	-122.381571
24	2511.885947	385.386363	-134.632903

Figure U6. A part of representative raw EIS data.

Make sure all you EIS test are using the same frequency range then **crate an excel sheet** as shown in Figure U7, where first column are names, second column are labels, each following column is one feature which is either real impedance or imaginary impedance at a frequency.
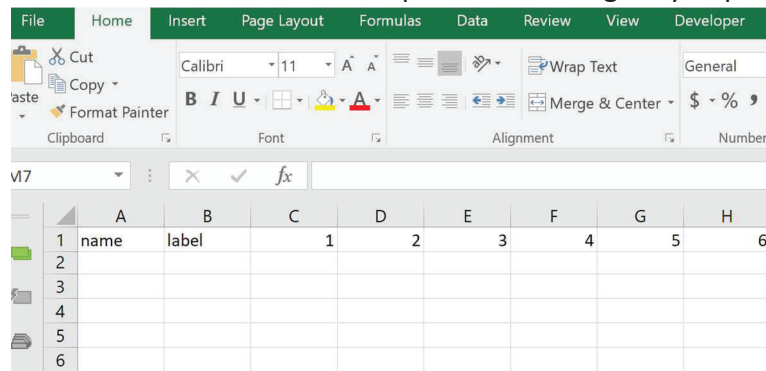
Figure U7. Excel template first step

Then, copy and transpose paste all real impedance (column B in Figure 12) to row 2 in Figure 13 starting at column C. (you may need to transpose paste the data in its original sheet then copy and paste to the template).

Next, using similar method as described above, copy all imaginary impedance (column C in Figure 12) to row 2 in Figure 13 starting at column CA (I have 76 frequencies in one EIS, I named row 1 from C as 1,2,3…76,77, to 152). Use your own naming system to give each EIS scan a unique name and put it in column A. Set a label (0 or 1 for binary classification) and put it in column B as shown in Figure U8.

A	B	C	BZ	CA	CB	EW	EX
1 name	label	1	76	77	78	151	152
2 4B2G1		0	258.054229	1621.41	-7.63751	-17.8826	-873.914	-939.174

Figure U8. Pasted imaginary impedance into row 2 from column CA to EX
(columns D-BY and CC-EV are hidden). Name and label added

Transfer all you EIS data that you are interested in classification (usually with similar experimental setting and a single variable) to the template and save it as *.csv file

Figure U9. Example raw EIS dataset ready for SVM machine learning

This example file can be downloaded from:
https://www.dropbox.com/s/7ftbm75mpyj6mc1/dataProject.csv?dl=0

Each row is a data from one run of a sample (one EIS/CV scan)
The label column is used for supervised classification/regression. It could be binary or quantitate depends on your applications.
For example:
In this dataset, "0" represent the blank/negative EIS results with no acetone in the samples; "1" represent the positive results with 5mM acetone concentration in the samples.
If you are testing glucose with 0, 1, 10 and 100 mM, you may label them as either "0,1,2,3" or "0,1,10,100" depending on doing classification or regression.
Each column followed is called a "feature" that in EIS is the real or imaginary impedance at a certain frequency.

3.2. SVM code for EIS data

The code can be downloaded from:
https://www.dropbox.com/s/7ftbm75mpyj6mc1/dataProject.csv?dl=0

Opened it in "Jupiter notebook" and put the datafile prepared as above in the same folder.
Run the code by cells.

用 "Jupyter notebook"

Figure U10. reading EIS dataset and data preprocessing

This cell read the EIS dateset from 'dataProject.csv'

The data was scaled using preprocessing.scale(X). And a principle component analysis (PCA) was performed to decompose the data into two dimensional.
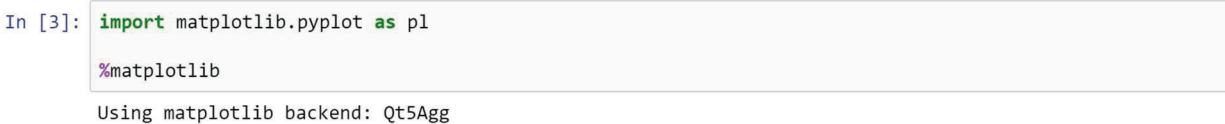
```
In [2]: #set training set
        from sklearn.model_selection import train_test_split
        X_train, X_test, y_train, y_test = \
                train_test_split(X_2d, y, test_size=0.20, random_state=1)
```

Figure U11. Split data into training and testing sets.

EIS data were split into 80% training and 20% testing.

```
In [3]: import matplotlib.pyplot as pl

        %matplotlib
        Using matplotlib backend: Qt5Agg
```

Figure U12. Pyplot from matplotlib package was imported and figures are set to appear in a new window.

Figures plotting tools was loaded.

```
In [4]:  #===========================================================================
         #Fig 2
         #===========================================================================
         h = .02  # step size in the mesh
         #C = 10  # SVM regularization parameter
         svc = svm.SVC(kernel='linear').fit(X_train, y_train)
         rbf_svc = svm.SVC(kernel='rbf').fit(X_train, y_train)
         poly_svc = svm.SVC(kernel='poly').fit(X_train, y_train)
         sigmod_svc = svm.SVC(kernel='sigmoid').fit(X_train, y_train)

         # create a mesh to plot in
         x_min, x_max = X_train[:, 0].min() - 1, X_train[:, 0].max() + 1
         y_min, y_max = X_train[:, 1].min() - 1, X_train[:, 1].max() + 1
         xx, yy = np.meshgrid(np.arange(x_min, x_max, h),
                              np.arange(y_min, y_max, h))

         # title for the plots
         titles = ['SVC with linear kernel',
                   'SVC with sigmoid kernel',
                   'SVC with RBF kernel',
                   'SVC with polynomial (degree 3) kernel']

         for k, clf in enumerate((svc, sigmod_svc, rbf_svc, poly_svc)):
             # Plot the decision boundary. For that, we will assign a color to each
             # point in the mesh [x_min, x_max]x[y_min, y_max].
             pl.subplot(2, 2, k + 1)
             pl.subplots_adjust(wspace=0.4, hspace=0.4)

             Z = clf.predict(np.c_[xx.ravel(), yy.ravel()])

             # Put the result into a color plot
             Z = Z.reshape(xx.shape)
             pl.contourf(xx, yy, Z, cmap=pl.cm.coolwarm, alpha=0.5)

             # Plot also the training points
             pl.scatter(X_2d[:, 0], X_2d[:, 1], c=y, cmap=pl.cm.coolwarm,s=20)

             #Plot score

             pl.text(0.6,0.9, ('Test accuracy: %.2f' % clf.score(X_2d, y)), transform=pl.gca().transAxes,
                     bbox={'facecolor':'white'})

             pl.xlabel('Sepal length')
             pl.ylabel('Sepal width')
             pl.xlim(xx.min(), xx.max())
             pl.ylim(yy.min(), yy.max())
             pl.xticks(())
             pl.yticks(())
             pl.title(titles[k])

         pl.show()
```

Figure U13. code cell for kernel selection
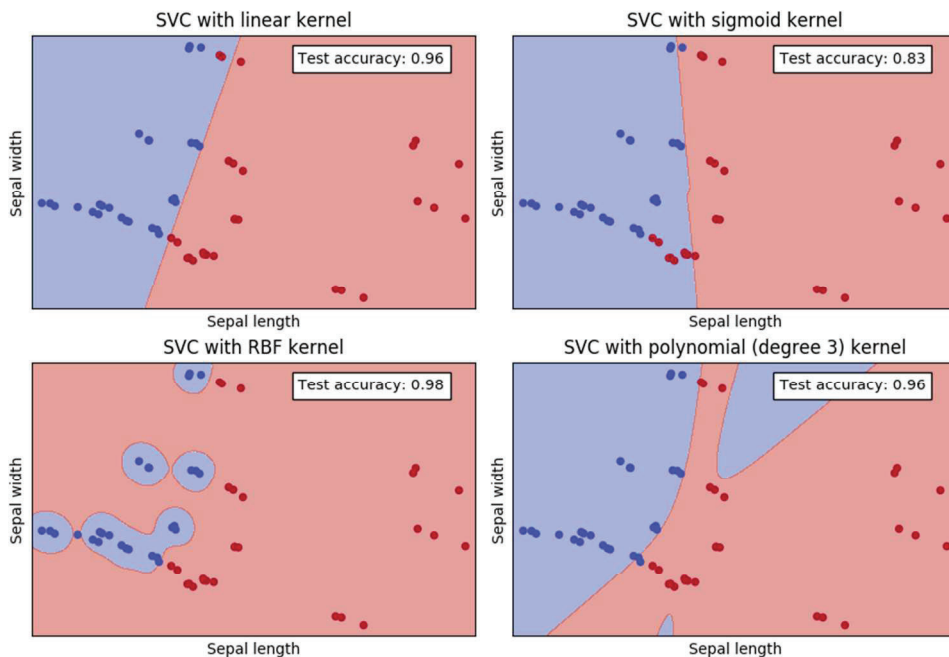
Figure U14. Result figure from the kernel selection code

```
In [6]:  from sklearn.model_selection import cross_val_score

         scores = cross_val_score(estimator=svm.SVC(kernel='rbf', gamma=0.01, C=10),
                                  X=X_2d,
                                  y=y,
                                  cv=10,
                                  n_jobs=10)
         print('CV accuracy scores: %s' % scores)
         print('CV accuracy: %.3f +/- %.3f' % (np.mean(scores), np.std(scores)))

         CV accuracy scores: [ 0.66666667  1.         0.83333333  1.         1.         1.         1.
           1.         1.         1.        ]
         CV accuracy: 0.950 +/- 0.107
```

Figure U15. Cross validation and its result.

```
In [7]:  from matplotlib.colors import Normalize

         from sklearn.svm import SVC
         from sklearn.preprocessing import StandardScaler
         from sklearn.datasets import load_iris
         from sklearn.model_selection import StratifiedShuffleSplit
         from sklearn.model_selection import GridSearchCV
         %matplotlib

         class MidpointNormalize(Normalize):

             def __init__(self, vmin=None, vmax=None, midpoint=None, clip=False):
                 self.midpoint = midpoint
                 Normalize.__init__(self, vmin, vmax, clip)

             def __call__(self, value, clip=None):
                 x, y = [self.vmin, self.midpoint, self.vmax], [0, 0.5, 1]
                 return np.ma.masked_array(np.interp(value, x, y))


         C_range = np.logspace(-5, 8, 14)
         gamma_range = np.logspace(-9, 4, 14)
         param_grid = dict(gamma=gamma_range, C=C_range)
         cv = StratifiedShuffleSplit(n_splits=10, test_size=0.2, random_state=12)
         grid = GridSearchCV(SVC(), param_grid=param_grid, cv=cv)
         grid.fit(X_train, y_train)

         print("The best parameters are %s with a score of %0.2f"
               % (grid.best_params_, grid.best_score_))

         # Now we need to fit a classifier for all parameters in the 2d version
         # (we use a smaller set of parameters here because it takes a while to train)

         C_2d_range = [1,10, 100, 1000,10000]
         gamma_2d_range = [1e-4,1e-3, 1e-2,1e-1,1]
         classifiers = []
         for C in C_2d_range:
             for gamma in gamma_2d_range:
                 clf = SVC(C=C, gamma=gamma)
                 clf.fit(X_train, y_train)
                 classifiers.append((C, gamma, clf))

         Using matplotlib backend: Qt5Agg
         The best parameters are {'C': 10.0, 'gamma': 0.01} with a score of 0.99
```

Figure U16. Finding the best parameters for RBF kernel and its results.

```
In [9]:  import matplotlib.pyplot as plt
         #plt.figure(figsize=(8, 6))
         #xx, yy = np.meshgrid(np.linspace(-3, 3, 200), np.linspace(-3, 3, 200))
         h=.1
         x_min, x_max = X_train[:, 0].min() - 1, X_train[:, 0].max() + 1
         y_min, y_max = X_train[:, 1].min() - 1, X_train[:, 1].max() + 1
         xx, yy = np.meshgrid(np.arange(x_min, x_max, h),
                              np.arange(y_min, y_max, h))
         for (k, (C, gamma, clf)) in enumerate(classifiers):
             # evaluate decision function in a grid
             Z = clf.decision_function(np.c_[xx.ravel(), yy.ravel()])
             Z = Z.reshape(xx.shape)

             # visualize decision function for these parameters
             plt.subplot(len(C_2d_range), len(gamma_2d_range), k + 1)
             plt.subplots_adjust(wspace=0.3, hspace=0.3)
             plt.title("gamma=10^%d, C=10^%d" % (np.log10(gamma), np.log10(C)),
                       size='medium')

             # visualize parameter's effect on decision function
             Z = clf.predict(np.c_[xx.ravel(), yy.ravel()])
             Z = Z.reshape(xx.shape)
             plt.contourf(xx, yy, Z, cmap=pl.cm.coolwarm, alpha=0.5)
             #plt.pcolormesh(xx, yy, -Z, cmap=plt.cm.RdBu)
             #plt.scatter(X_train[:, 0], X_train[:, 1], c=y_train, cmap=pl.cm.coolwarm)
             #plt.scatter(X_test[:, 0], X_test[:, 1], y_test, cmap=pl.cm.RdBu_r)

             for i in range(0, X_train.shape[0]):
                 if y_train[i] == 0:
                     c1 = pl.scatter(X_train[i,0],X_train[i,1],c='r',marker='o', s = 20 )
                 elif y_train[i] == 1:
                     c2 = pl.scatter(X_train[i,0],X_train[i,1],c='g',marker='+', s = 30 )
             for i in range(0, X_test.shape[0]):
                 if y_test[i] == 0:
                     c3 = pl.scatter(X_test[i,0],X_test[i,1],c='b',marker='o', s = 20 )
                 elif y_test[i] == 1:
                     c4 = pl.scatter(X_test[i,0],X_test[i,1],c='m',marker='+', s = 30 )


             plt.xticks(())
             plt.yticks(())
             plt.axis('tight')

         scores = grid.cv_results_['mean_test_score'].reshape(len(C_range),
                                               len(gamma_range))
```

Figure U17. Code to visualize effect of parameters for RBF kernel.
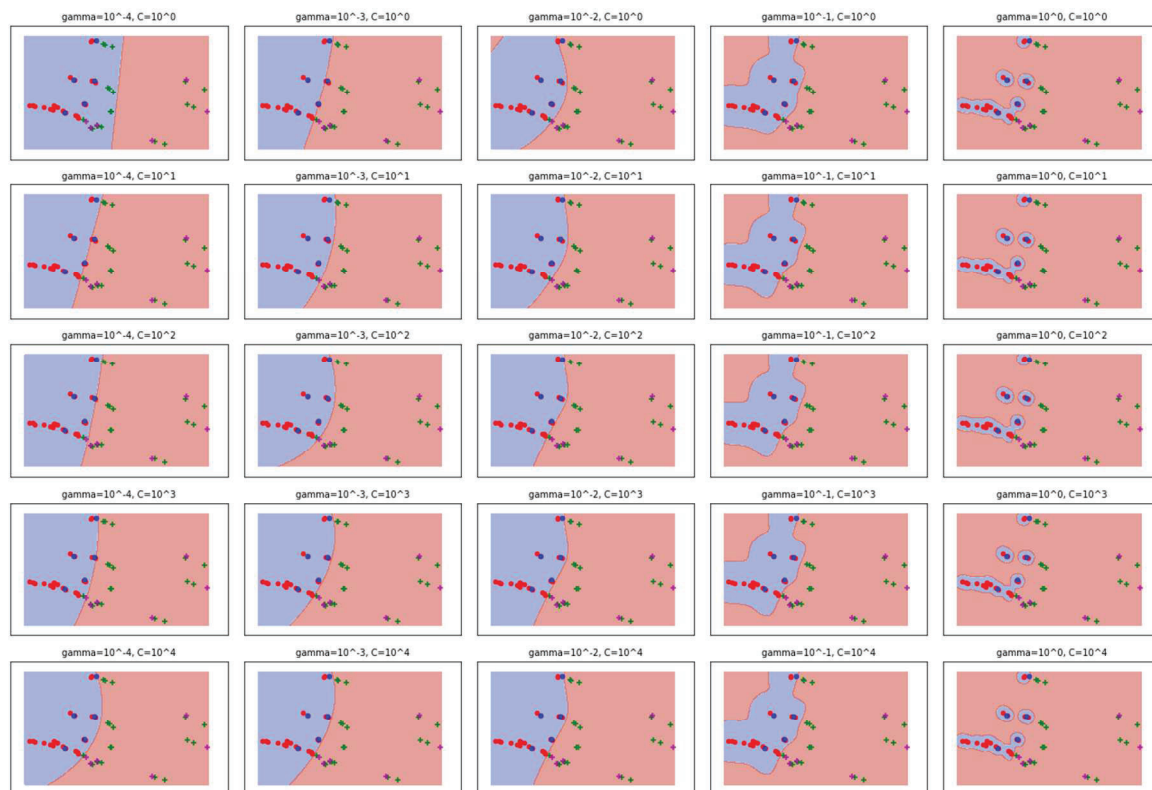
Figure U18. Effect of parameters for RBF kernel.

If you have your dataset prepared like this you may easily modify my code or the codes available online and test it out.

3.3. Scikit-Learn

http://scikit-learn.org/stable/index.html

If you have time, checkout this site.

Most of my code and knowledge is based on the information and example from this open source project.

Good Luck!

10/23/2017

SVM Machine Learning for EIS data analysis step-by-step user guide

支持向量机机器学习用于电化学阻抗谱数据分析的使用指南

Yue Rong

荣越

UF Biosensor Lab

佛罗里达大学生物传感器实验室

1. Software preparation（软件准备）

1.1. Download Anaconda installer（下载 anaconda 安装包）

"Anaconda is a freemium open source distribution of the Python and R programming languages for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment." --https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)

"Anaconda 是一个免费开源的 Python 及 R 语言开发环境。它主要用于大规模的数据处理，预测分析和科学计算，致力于简化程序包的管理和使用"—来自维基百科

It includes all the packages and tools we need and is completely free.

该软件包含了我们全部需要的工具

It can be downloaded at:

可以在以下地址下载：
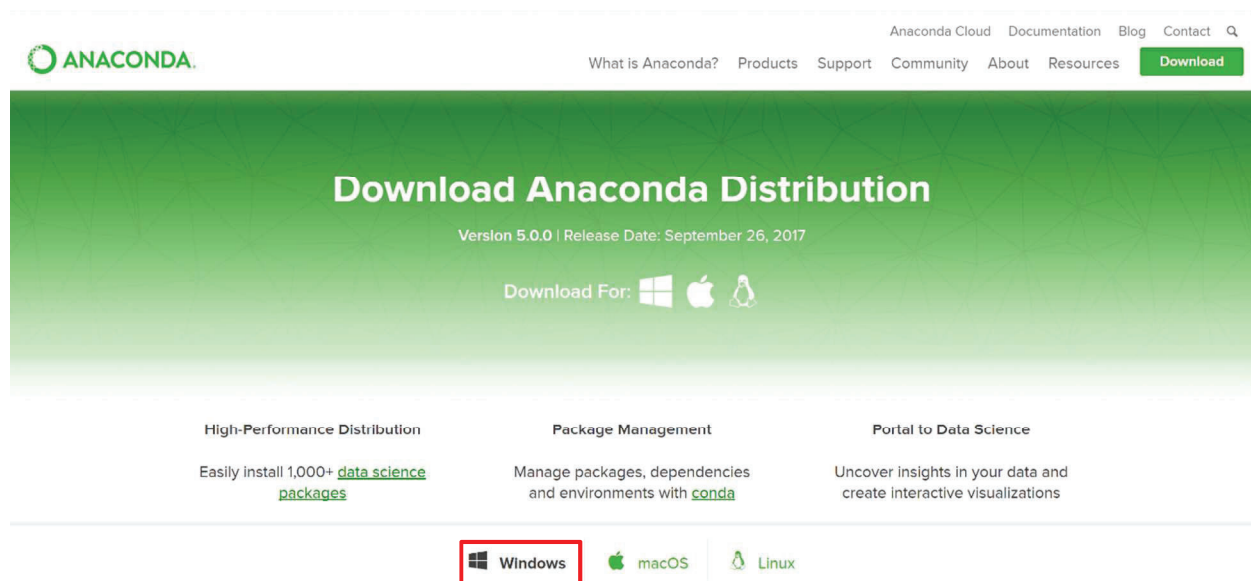
Figure 1 Choose your operating system by clicking "Windows", "macOS" or "Linux".

https://www.continuum.io/downloads

There are different versions of Anaconda for Windows, macOS and Linux; Python 3.6 and 2.7. It will be about 2 Gb after installation and downloading all packages. We suggest Python 3.6.

该网站提供了 Anaconda 为 Windows，macOS 和 Linux 三种平台的安装包。并且都有 Python 3.6 和 2.7 两个版本。下载安装完成大概需要 2Gb 的硬盘空间。我们建议使用 Python 3.6 版本。
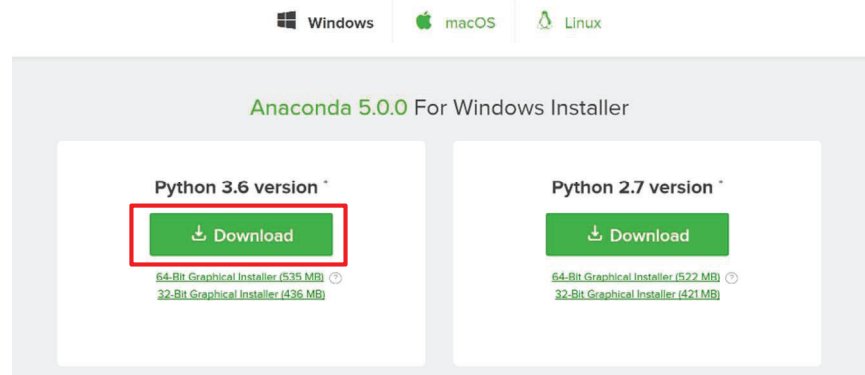
Figure 2 Click to download Python 3.6 version installer (macOS and Linux are similar)

1.2. Install Anaconda（安装 Anaconda）

Double click the installer to launch.（双击安装包开始）

NOTE: If you encounter any issues during installation, temporarily disable your anti-virus software during install, then re-enable it after the installation concludes. If you have installed for all users, uninstall Anaconda and re-install it for your user only and try again.

注意：如果在安装过程中遇到任何问题，暂时关闭反病毒软件，安装完成后重新开启。如果你选择了"为所有用户"安装，请删除 Anaconda，重新安装并选择"只为自己"安装。

Click Next.（点击继续）

Read the licensing terms and click "I Agree".（阅读并点击同意）

Select an install for "Just Me" unless you're installing for all users (which requires Windows Administrator privileges) and click Next.

选择"只为我"安装，除非你要为所有用户安装（要求 Windows 管理员权限）点击继续。

Select a destination folder to install Anaconda and click the Next button.（选择安装位置）

NOTE: Install Anaconda to a directory path that does not contain spaces or unicode characters.

注意：安装目录中不要含有空格或者特殊符号

NOTE: Do not install as Administrator unless admin privileges are required.

注意：不要以管理员身份安装，除非管理员权限需要的时候
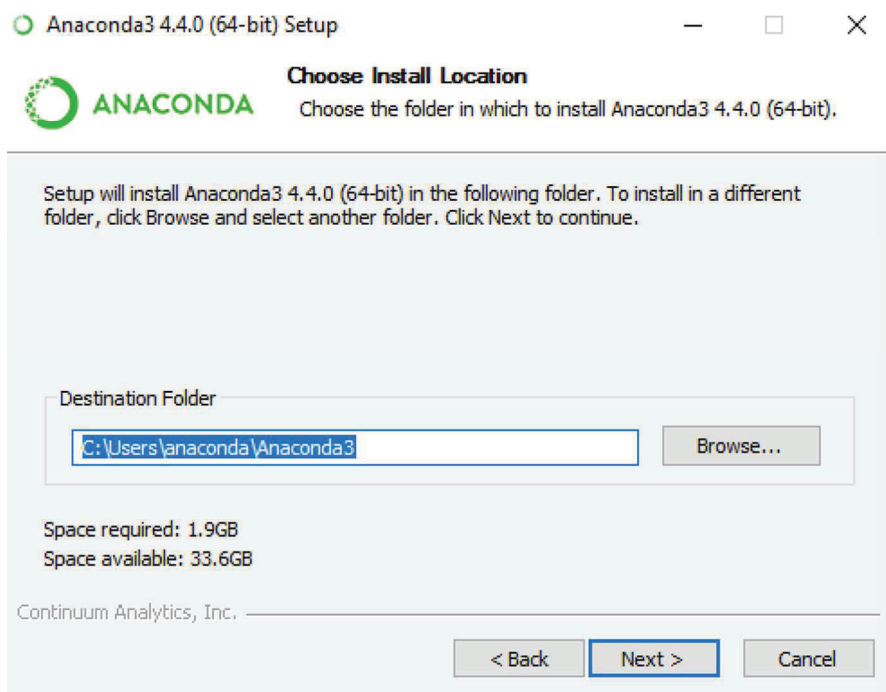
Figure 3 Select a destination folder to install Anaconda and click the Next button

Choose whether to add Anaconda to your PATH environment variable.（选择是否将 Anaconda 加入你的 PATH 环节变量）

We recommend not adding Anaconda to the PATH environment variable, since this can interfere with other software. Instead, use Anaconda software by opening Anaconda Navigator or the Anaconda Prompt from the Start Menu.

我们不推荐把 Anaconda 加入 PATH 环境变量，因为这可能会影响其他软件。所以我们建议从开始菜单打开 Anaconda Navigator 或者 Anaconda Prompt 来使用。
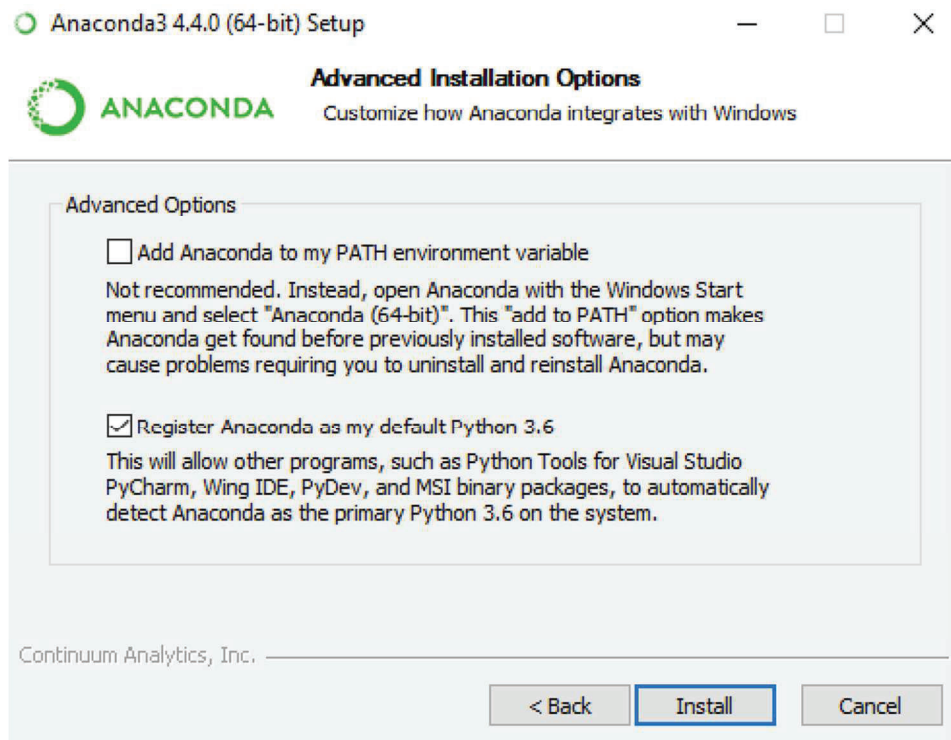
Figure 4 register Anaconda as your default Python 3.6

Choose whether to register Anaconda as your default Python 3.6. （选择是否把 Anaconda 设为默认的 Python 3.6）

Unless you plan on installing and running multiple versions of Anaconda, or multiple versions of Python, you should accept the default and leave this box checked.

除非你有多版本的 Anaconda 或者多版本的 Python，你应该接受这个默认设置。

Click the Install button. （点击安装）

You can click Show Details if you want to see all the packages Anaconda is installing.

你可以点击产看详情，如果你想看到所有 Anaconda 正在安装的程序包

Click the Next button. （点击继续）

After a successful installation, you will see the "Thanks for installing Anaconda" dialog box:

安装完成后，你会看到"感谢安装 Anaconda"的对话框
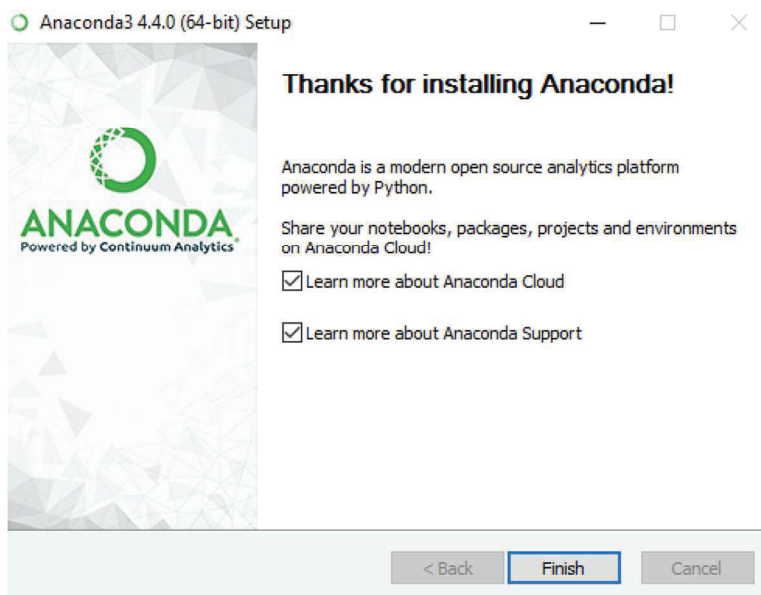
Figure 5 successful installation dialog

Click the Finish button.（点击完成）

After your install is complete, verify it by opening Anaconda Navigator, a program that is included with Anaconda: from your Windows Start menu, select the shortcut Anaconda Navigator. If Navigator opens, you have successfully installed Anaconda.

安装完成后，通过打开 Anaconda Navigator 来确认。如果你能打开 Anaconda Navigator，说明你已经成功安装了 Anaconda。

If you had any issue installing the Anaconda, please visit https://docs.anaconda.com/anaconda/install/ for help.

如果你在安装过程中遇到其他问题，请访问：https://docs.anaconda.com/anaconda/install/

2. Getting started（准备开始）
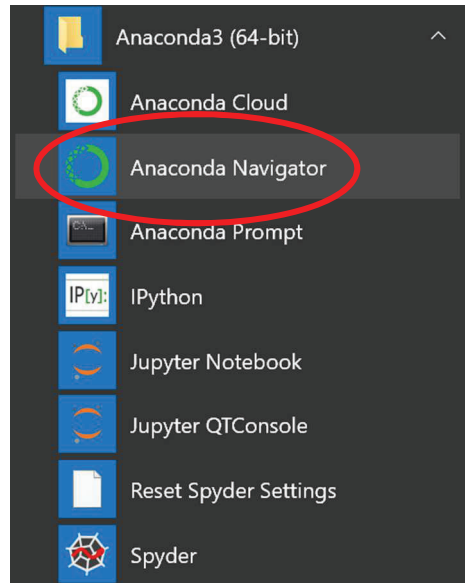
Figure 6 Find Anaconda 3 and open "Anaconda Navigator"

2.1. Open Anaconda Navigator（打开 Anaconda Navigator）

Find Anaconda 3 in "Windows Start Menu" and click "Anaconda Navigator" to open
在开始菜单中找到 Anaconda 3 并双击"Anaconda Navigator"打开
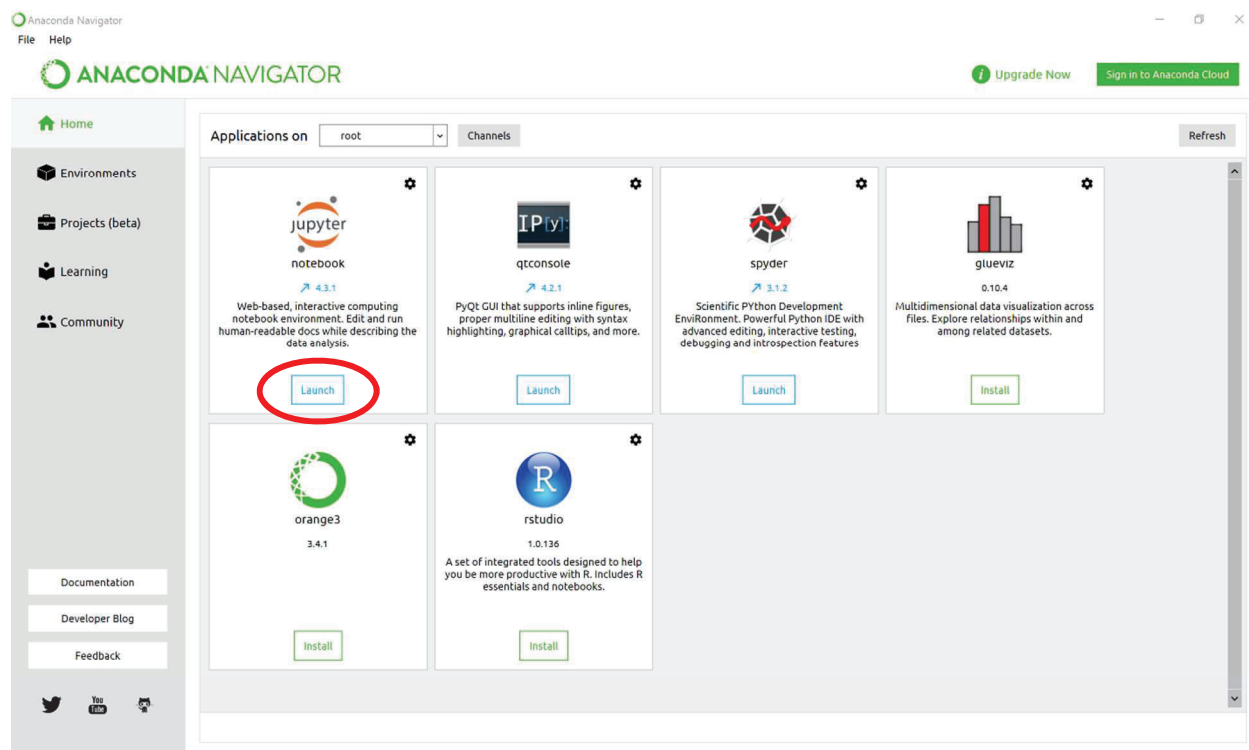
2.2. Click "Launch" to open "Jupyter Notebook"（点击"Jupyter notebook"下的"Launch"按钮）

It will open in your internet browser (for me it's a Chrome) an the default file location is under your
system user folder. (for example C:\Users\Username)

Figure 7 Lunch "Jupyter Notebook" in Anaconda Navigator

它会在你默认的网页浏览器（我用的是 Chrome）中打开，默认的文件位置是在你的系统用户文件夹。
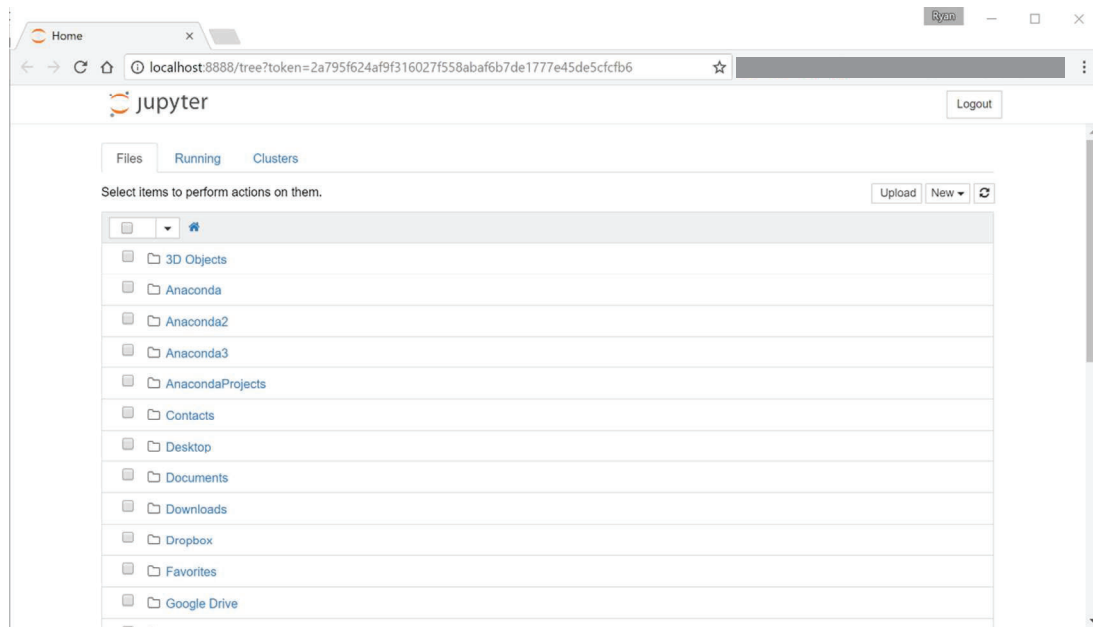
Figure 8 "Jupyter Notebook" dashboard

2.3. Run an example code（运行一个测试代码）
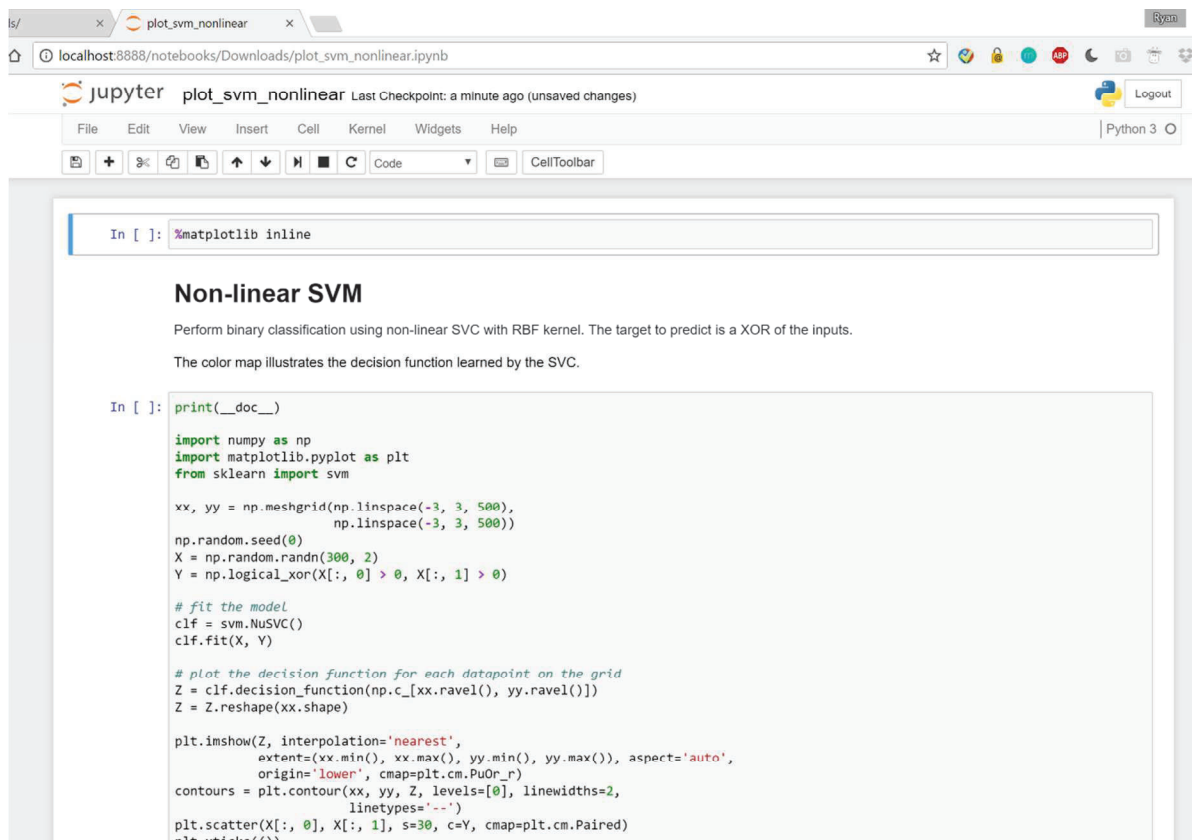
Download an example code from the link below and save it to "downloads" folder.

从以下地址下载一个测试代码，并保存到"下载"目录

http://scikit-learn.org/stable/_downloads/plot_svm_nonlinear.ipynb

use the file browser to navigate and open the download source code above

通过网页内的文件浏览器找到并打开刚下载的测试代码

Figure 9 Opened example code in "Jupyter Notebook"

It should look like the screenshot above, and in "Cell" click "Run All" or run cells by clicking the [▶] .
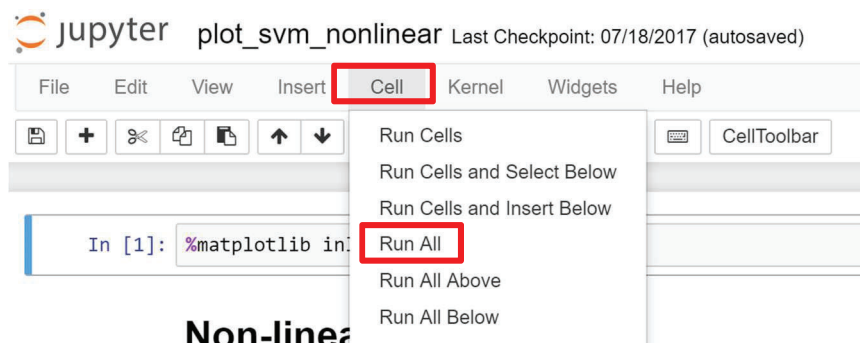
Figure 10 Run All cells to test the example code

打开后应该如上图的截屏一样。在"Cell"里点击"Run All"或者单个段落运行。

If it generated a figure like this, your software is good to go. Cheers!

如果它生成了下图，那么你的软件就没问题了！

```
plt.scatter(X[:, 0], X[:, 1], s=30, c=Y, cmap=plt.cm.Paired)
plt.xticks(())
plt.yticks(())
plt.axis([-3, 3, -3, 3])
plt.show()
```

Automatically created module for IPython interactive environment
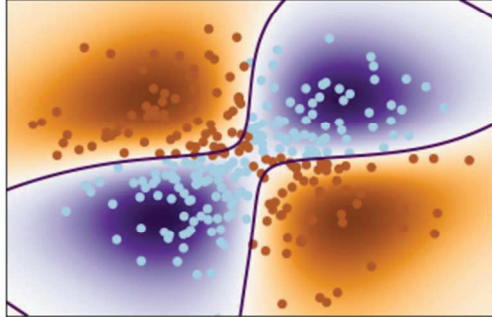
Figure 11 Expected output from example code

3. EIS data analysis with SVM machine learning （用支持向量机机器学习对电化学阻抗谱数据进行分析）

3.1. Data preparation（数据准备）

Raw EIS data usually includes frequency with real impedance (Z') and imaginary impedance (Z'') at that frequency.

电化学阻抗谱的原始数据通常由在特定频率的实阻抗和虚阻抗构成。

	A	B	C
	Frequency [Hz]	Z'_raw [Ohm]	Z''_raw [Ohm]
0	99999.99984	258.054229	-7.637512
1	85769.58917	255.299691	-17.882566
2	73564.22411	278.492955	-18.88002
3	63095.73291	275.258838	-15.963164
4	54116.95083	289.347566	-28.062849
5	46415.88639	287.247927	-26.796648
6	39810.71513	286.058176	-32.585536
7	34145.48673	296.438676	-33.791812
8	29286.44376	301.039143	-34.728174
9	25118.8625	300.781714	-35.632715
10	21544.34515	307.305851	-40.781186
11	18478.4963	308.219676	-39.220288
12	15848.93039	312.550444	-44.296644
13	13593.56249	311.666707	-50.051774
14	11659.14268	315.616756	-54.547519
15	9999.998792	318.54106	-58.196081
16	8576.957885	322.457927	-63.004504
17	7356.421543	326.558636	-69.321829
18	6309.57253	330.949427	-77.043239
19	5411.69444	336.52842	-84.527607
20	4641.588089	343.795531	-92.286639
21	3981.071029	352.501566	-102.064143
22	3414.548269	359.595675	-112.278786
23	2928.644022	372.104547	-122.381571
24	2511.885947	385.386363	-134.632903

Figure 12 A part of representative raw EIS data.

Make sure all you EIS test are using the same frequency range then **crate an excel sheet** as shown below, where first column are names, second column are labels, each following column is one feature which is either real impedance or imaginary impedance at a frequency.

确定你所有电化学阻抗谱的测试使用的都是相同的频率分布，并按下图建一个新的 excel 表格。

Figure 13 Excel template first step

第一列是名称，第二列是标签，之后的每一列是实阻抗护着虚阻抗。

Then, copy and transpose paste all real impedance (column B in Figure 12) to row 2 in Figure 13 starting at column C. (you may need to transpose paste the data in its original sheet then copy and paste to the template)

然后，把所有图 12 中 B 列的实阻抗数据，转置粘贴到图 13 中第二行，从第 C 列开始的位置
（在原表格中先转置再复制粘贴可能更容易）
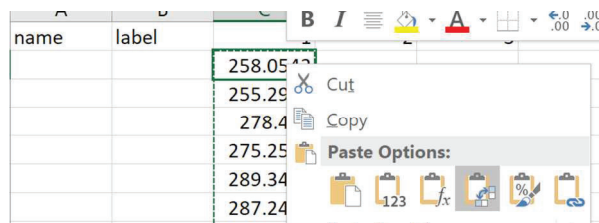
Figure 14 Using transpose paste option to paste all real impedance data in a row.

Then using similar method to copy all imaginary impedance (column C in Figure 12) to row 2 in Figure 13 starting at column CA (I have 76 frequencies in one EIS, I named row 1 from C as 1,2,3...76,77, to 152). Use your own naming system to give each EIS scan a unique name and put it in column A. Set a label (0 or 1 for binary classification) and put it in column B.

然后，用类似的方法，把所有虚阻抗的数据（图 12 中 C 列的数据）转置粘贴到图 13 中，第二行，CA 列开始的位置（在我的电化学阻抗谱中用了 76 个频率，所以第一行从 C 开始命名为 1,2,3 到 76，然后 77,78 一直到 152）。你可以使用你自己的命名系统，在 A 列中给每一个数据列命名，然后在 B 列中给每一个数据设置标签（二进制分类中就是 0 或者 1）

	A	B	C	BZ	CA	CB	EW	EX
1	name	label	1	76	77	78	151	152
2	4B2G1	0	258.054229	1621.41	-7.63751	-17.8826	-873.914	-939.174

Figure 15 Pasted imaginary impedance into row 2 from column CA to EX (columns D-BY and CC-EV are hidden). Name and label added

Transfer all you EIS data that you are interested in classification (usually with similar experimental setting and a single variable) to the template and save it as *.csv file

把那些你所有想要分类的的电化学阻抗谱数据（通常是相似的单变量实验）都用同样的方式粘贴到这个模板中，并另存为 csv 格式。

name	Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4B2G1	0	258.0542	255.2997	278.493	275.2588	289.3476	287.2479	286.0582	296.4387	301.0391	300.7817	307.3059	308.2197	312.5504	311.6667	315.6168	318.5411
4B2G2	0	265.8894	268.0702	273.757	285.1902	278.4257	290.3433	291.5499	293.6509	302.0073	302.9377	307.6231	308.7262	312.3669	313.9265	315.5427	319.354
4B2G3	0	267.317	266.4577	276.3333	275.4226	286.8759	282.8898	294.9838	298.7177	299.2943	304.1186	306.9148	307.7466	312.0538	312.9816	315.566	318.3924
5B2G1	0	212.9031	210.1753	226.3933	227.7355	227.9052	226.8304	234.8648	238.0592	239.7823	242.3075	242.45	245.1441	246.9025	247.1612	249.6174	253.1013
5B2G2	0	221.8998	225.2367	231.5692	227.6233	227.1677	228.3944	234.2414	240.8749	242.4335	243.288	244.2544	247.5539	247.3955	248.675	250.9991	253.9048
5B2G3	0	220.5023	213.8915	224.6013	228.9741	225.4477	231.434	239.4562	239.4411	240.5017	244.0199	246.1277	246.6161	248.3261	249.4041	251.0594	253.8992
6B2G1	0	166.9949	173.7204	184.3192	169.5682	173.2099	176.5317	179.8662	179.617	183.9284	185.823	186.4487	188.3658	190.5809	191.4963	193.6587	196.2139
6B2G2	0	179.0694	182.0632	171.9485	169.2709	183.162	179.8637	183.721	182.0691	184.8741	185.6762	187.9043	189.6923	191.1944	191.7999	192.7237	196.459
6B2G3	0	168.6349	182.5261	178.646	174.2017	177.7401	179.4154	182.8734	184.3721	187.3621	184.121	186.3112	189.6721	191.2342	191.5679	193.5646	196.9216
4B2G5A1	1	259.9313	259.0188	266.8876	284.1592	285.0439	286.6123	294.9715	297.3411	304.1833	302.6753	305.6459	309.1808	311.1617	310.8451	313.0142	317.0626
4B2G5A2	1	263.466	266.4305	283.5566	278.298	285.1393	285.0858	289.288	298.2087	300.9645	302.7892	303.546	306.7226	309.8123	310.6402	311.6402	316.1097
4B2G5A3	1	265.7502	266.4175	272.3123	276.0552	276.7814	288.0103	291.8241	293.7257	299.5611	301.3291	303.3106	303.8075	308.4623	310.1837	312.8859	316.5454
5B2G5A1	1	214.0163	227.5339	223.9881	230.3015	225.7774	232.9494	236.9239	240.9703	242.9936	241.4223	245.6318	249.8173	249.9303	249.7395	251.7926	253.9502
5B2G5A2	1	221.6863	218.2461	224.2944	225.4269	234.8084	235.1881	237.1379	239.1588	242.6543	243.4588	245.48	247.0395	249.6863	250.0278	251.7821	254.6247
5B2G5A3	1	219.98	212.6655	217.8348	232.9979	229.0809	234.607	239.0655	242.7952	240.4469	246.9357	247.8388	248.3381	250.0531	249.12	250.7349	254.3541
6B2G5A1	1	186.333	178.2627	177.4258	185.1232	182.1481	184.8892	182.5009	183.927	188.007	190.4824	191.2653	190.0282	193.1123	193.4678	194.8699	196.8375
6B2G5A2	1	179.8914	173.6595	173.5661	180.1657	180.2534	182.6176	184.5293	184.7471	186.4101	187.6895	187.2631	189.8662	190.9922	191.2556	192.9192	194.9413
6B2G5A3	1	182.3292	179.924	177.2412	167.8647	180.4624	179.4651	185.4587	183.6399	188.3154	186.6553	189.636	191.809	191.5059	193.1888	195.5724	
4B1G5A1	1	286.9791	291.5914	297.6871	292.363	306.4467	314.4559	308.4549	322.7986	324.8604	325.3788	325.2187	333.2716	331.9342	333.5523	336.6442	337.86
4B1G5A2	1	272.9672	281.8878	288.8579	313.4164	310.8411	312.8765	312.6907	309.6892	323.1059	325.5448	327.3001	327.981	332.406	333.0716	333.1723	337.3366
4B1G5A3	1	295.2762	283.1599	300.0213	297.0929	299.5606	307.4459	310.4039	321.0305	319.0035	325.234	326.0285	330.9777	332.898	331.6771	334.3737	338.3772

Figure 16 Example raw EIS dataset ready for SVM machine learning

This example file can be downloaded from:

我的实例文件可以从以下地址下载（国内可能需要翻墙）

https://www.dropbox.com/s/7ftbm75mpyj6mc1/dataProject.csv?dl=0

Each row is a data from one run of a sample (one EIS/CV scan)

The label column is used for supervised classification/regression. It could be binary or quantitate depends on your applications.

每一行代表一个完整的电化学阻抗谱。标签列用于有监督的分类或者回归。它可以是二进制的也可以是定量的。

For example:

In this dataset, "0" represent the blank/negative EIS results with no acetone in the samples; "1" represent the positive results with 5mM acetone concentration in the samples.

比如说：

在这个数据中，0 代表了空白对照，1 代表了有 5mM 丙酮的样本。

If you are testing glucose with 0, 1, 10 and 100 mM, you may label them as either "0,1,2,3" or "0,1,10,100" depending on doing classification or regression.

如果你要检测葡萄糖在 0,1,10 和 100mM 这几个浓度，那你可以给对应数据加上"0,1,2,3"或者"0,1,10,100"的标签。取决于你是要分类分析还是回归分析。

Each column followed is called a "feature" that in EIS is the real or imaginary impedance at a certain frequency.

之后每一列都是一个机器学习中的"特征"，它们都是电化学阻抗谱中某一频率下的实阻抗或者虚阻抗

3.2. SVM code for EIS data （支持向量机用于电化学阻抗谱的代码）

The code can be downloaded from:
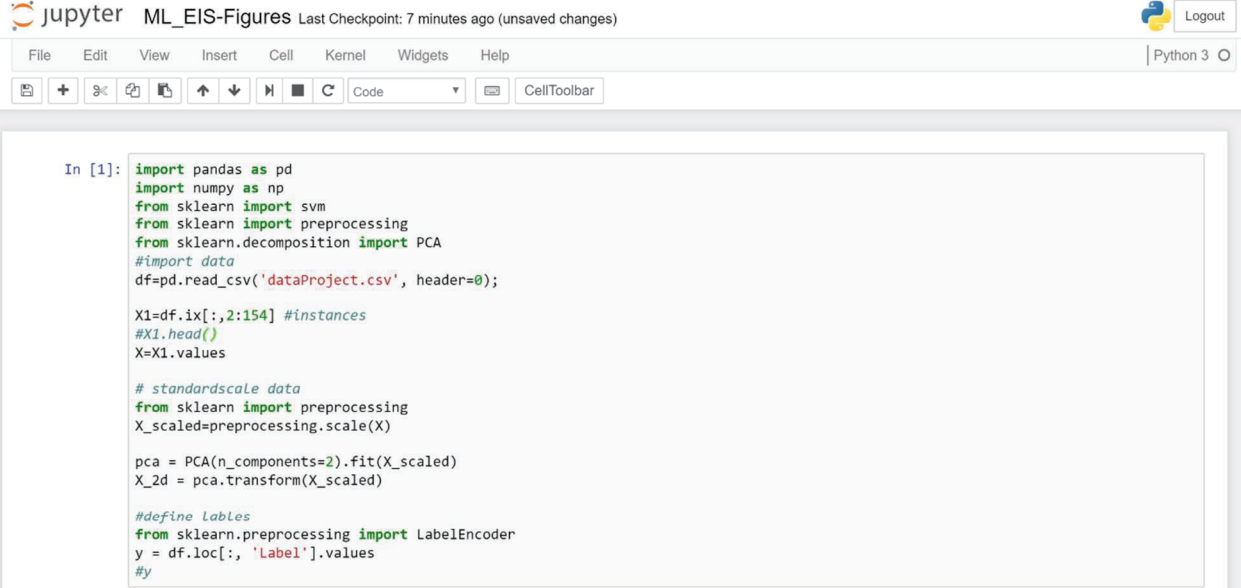
代码可以从以下地址下载（国内可能需要翻墙）

https://www.dropbox.com/s/7ftbm75mpyj6mc1/dataProject.csv?dl=0

Opened it in "Jupiter notebook" and put the datafile prepared as above in the same folder.
Run the code by cells.
用"Jupyter notebook"打开，并把之前的数据文件放到相同的目录。

Figure 17 reading EIS dataset and data preprocessing

This cell read the EIS dateset from 'dataProject.csv'
The data was scaled using preprocessing.scale(X). And a principle component analysis (PCA) was performed to decompose the data into two dimensional.
这一个段落作用是从"dataProject.csv"读取数据。数据之后通过 preprocessing.scale(X)标准化，通过主成分分析实现降维。

Figure 18 Split data into training and testing sets.

EIS data were split into 80% training and 20% testing.
电化学阻抗谱数据组被分为 80%的训练组和 20%的检验组

Figure 19 pyplot from matplotlib package was imported and figures are set to appear in a new window.

Figures plotting tools was loaded. 制图工具载入

Figure 21 result figure from the kernel selection code 核选择代码的结果图

```
In [4]:  #===========================================================================
         #Fig 2
         #===========================================================================
         h = .02  # step size in the mesh
         #C = 10  # SVM regularization parameter
         svc = svm.SVC(kernel='linear').fit(X_train, y_train)
         rbf_svc = svm.SVC(kernel='rbf').fit(X_train, y_train)
         poly_svc = svm.SVC(kernel='poly').fit(X_train, y_train)
         sigmod_svc = svm.SVC(kernel='sigmoid').fit(X_train, y_train)

         # create a mesh to plot in
         x_min, x_max = X_train[:, 0].min() - 1, X_train[:, 0].max() + 1
         y_min, y_max = X_train[:, 1].min() - 1, X_train[:, 1].max() + 1
         xx, yy = np.meshgrid(np.arange(x_min, x_max, h),
                              np.arange(y_min, y_max, h))

         # title for the plots
         titles = ['SVC with linear kernel',
                   'SVC with sigmoid kernel',
                   'SVC with RBF kernel',
                   'SVC with polynomial (degree 3) kernel']


         for k, clf in enumerate((svc, sigmod_svc, rbf_svc, poly_svc)):
             # Plot the decision boundary. For that, we will assign a color to each
             # point in the mesh [x_min, x_max]x[y_min, y_max].
             pl.subplot(2, 2, k + 1)
             pl.subplots_adjust(wspace=0.4, hspace=0.4)

             Z = clf.predict(np.c_[xx.ravel(), yy.ravel()])

             # Put the result into a color plot
             Z = Z.reshape(xx.shape)
             pl.contourf(xx, yy, Z, cmap=pl.cm.coolwarm, alpha=0.5)

             # Plot also the training points
             pl.scatter(X_2d[:, 0], X_2d[:, 1], c=y, cmap=pl.cm.coolwarm,s=20)

             #Plot score

             pl.text(0.6,0.9, ('Test accuracy: %.2f' % clf.score(X_2d, y)), transform=pl.gca().transAxes,
                     bbox={'facecolor':'white'})


             pl.xlabel('Sepal length')
             pl.ylabel('Sepal width')
             pl.xlim(xx.min(), xx.max())
             pl.ylim(yy.min(), yy.max())
             pl.xticks(())
             pl.yticks(())
             pl.title(titles[k])

         pl.show()
```

Figure 20 code cell for kernel selection 核选择的代码

```python
In [7]: from matplotlib.colors import Normalize

        from sklearn.svm import SVC
        from sklearn.preprocessing import StandardScaler
        from sklearn.datasets import load_iris
        from sklearn.model_selection import StratifiedShuffleSplit
        from sklearn.model_selection import GridSearchCV
        %matplotlib

        class MidpointNormalize(Normalize):

            def __init__(self, vmin=None, vmax=None, midpoint=None, clip=False):
                self.midpoint = midpoint
                Normalize.__init__(self, vmin, vmax, clip)

            def __call__(self, value, clip=None):
                x, y = [self.vmin, self.midpoint, self.vmax], [0, 0.5, 1]
                return np.ma.masked_array(np.interp(value, x, y))


        C_range = np.logspace(-5, 8, 14)
        gamma_range = np.logspace(-9, 4, 14)
        param_grid = dict(gamma=gamma_range, C=C_range)
        cv = StratifiedShuffleSplit(n_splits=10, test_size=0.2, random_state=12)
        grid = GridSearchCV(SVC(), param_grid=param_grid, cv=cv)
        grid.fit(X_train, y_train)

        print("The best parameters are %s with a score of %0.2f"
              % (grid.best_params_, grid.best_score_))

        # Now we need to fit a classifier for all parameters in the 2d version
        # (we use a smaller set of parameters here because it takes a while to train)

        C_2d_range = [1,10, 100, 1000,10000]
        gamma_2d_range = [1e-4,1e-3, 1e-2,1e-1,1]
        classifiers = []
        for C in C_2d_range:
            for gamma in gamma_2d_range:
                clf = SVC(C=C, gamma=gamma)
                clf.fit(X_train, y_train)
                classifiers.append((C, gamma, clf))

Using matplotlib backend: Qt5Agg
The best parameters are {'C': 10.0, 'gamma': 0.01} with a score of 0.99
```

Figure 23 Finding the best parameters for RBF kernel and its results 径向基函数核最佳参数选择的代码核

```python
In [6]: from sklearn.model_selection import cross_val_score

        scores = cross_val_score(estimator=svm.SVC(kernel='rbf', gamma=0.01, C=10),
                                 X=X_2d,
                                 y=y,
                                 cv=10,
                                 n_jobs=10)
        print('CV accuracy scores: %s' % scores)
        print('CV accuracy: %.3f +/- %.3f' % (np.mean(scores), np.std(scores)))

CV accuracy scores: [ 0.66666667  1.          0.83333333  1.          1.          1.          1.
  1.          1.          1.        ]
CV accuracy: 0.950 +/- 0.107
```

Figure 22 Cross validation and its result. 交叉验证的代码核结果

Figure 25 Effect of parameters for RBF kernel. 径向基函数核参数改变的效果

```python
In [9]: import matplotlib.pyplot as plt
        #plt.figure(figsize=(8, 6))
        #xx, yy = np.meshgrid(np.linspace(-3, 3, 200), np.linspace(-3, 3, 200))
        h=.1
        x_min, x_max = X_train[:, 0].min() - 1, X_train[:, 0].max() + 1
        y_min, y_max = X_train[:, 1].min() - 1, X_train[:, 1].max() + 1
        xx, yy = np.meshgrid(np.arange(x_min, x_max, h),
                             np.arange(y_min, y_max, h))
        for (k, (C, gamma, clf)) in enumerate(classifiers):
            # evaluate decision function in a grid
            Z = clf.decision_function(np.c_[xx.ravel(), yy.ravel()])
            Z = Z.reshape(xx.shape)

            # visualize decision function for these parameters
            plt.subplot(len(C_2d_range), len(gamma_2d_range), k + 1)
            plt.subplots_adjust(wspace=0.3, hspace=0.3)
            plt.title("gamma=10^%d, C=10^%d" % (np.log10(gamma), np.log10(C)),
                      size='medium')

            # visualize parameter's effect on decision function
            Z = clf.predict(np.c_[xx.ravel(), yy.ravel()])
            Z = Z.reshape(xx.shape)
            plt.contourf(xx, yy, Z, cmap=pl.cm.coolwarm, alpha=0.5)
            #plt.pcolormesh(xx, yy, -Z, cmap=plt.cm.RdBu)
            #plt.scatter(X_train[:, 0], X_train[:, 1], c=y_train, cmap=pl.cm.coolwarm)
            #plt.scatter(X_test[:, 0], X_test[:, 1], y_test, cmap=pl.cm.RdBu_r)

            for i in range(0, X_train.shape[0]):
                if y_train[i] == 0:
                    c1 = pl.scatter(X_train[i,0],X_train[i,1],c='r',marker='o', s = 20 )
                elif y_train[i] == 1:
                    c2 = pl.scatter(X_train[i,0],X_train[i,1],c='g',marker='+', s = 30 )
            for i in range(0, X_test.shape[0]):
                if y_test[i] == 0:
                    c3 = pl.scatter(X_test[i,0],X_test[i,1],c='b',marker='o', s = 20 )
                elif y_test[i] == 1:
                    c4 = pl.scatter(X_test[i,0],X_test[i,1],c='m',marker='+', s = 30 )


            plt.xticks(())
            plt.yticks(())
            plt.axis('tight')

        scores = grid.cv_results_['mean_test_score'].reshape(len(C_range),
                                                             len(gamma_range))
```

Figure 24 Code to visualize effect of parameters for RBF kernel. 径向基函数核参数效果展示的代码

If you have your dataset prepared like this you may easily modify my code or the codes available online and test it out.

如果你的数据用这种方式预处理好了，那么你可以很容易的通过修改我的代码或者使用网上可用的代码来进行测试

3.3. Scikit-Learn

http://scikit-learn.org/stable/index.html

If you have time, checkout this site.

Most of my code and knowledge is based on the information and example from this open source project.

如果你有时间，请查阅以上网址，我的所有代码都是基于 Scikit-learn 这个开源项目。

Good Luck! 祝你好运！

10/23/2017

SNAPS
PART-1 & PART-2

Shoumen Datta

HAPHAZARD REALITY – IOT IS A METAPHOR

Sensor Analytics Point Solutions (SNAPS) for Detection and Decision Support Systems (DSS)

Eric S. McLamore[1*], Shoumen Palit Austin Datta[1, 2, 3 **], Victoria Morgan[1], Nicholas Cavallaro[1], Greg Kiker[1], Daniel M. Jenkins[4], Yue Rong[1], Carmen Gomes[5], Jonathan Claussen[5], Diana Vanegas[6], Sahar Bahramzadeh[7], Evangelyn C. Alocilja[8, 9]

[1] Agricultural and Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL 32611

[2] MIT Auto-ID Labs, Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

[3] MDPnP Labs, Biomedical Engineering Program, Department of Anesthesiology, Massachusetts General Hospital, Harvard Medical School, 65 Landsdowne Street, Cambridge, MA 02139

[4] Molecular Biosciences and Bioengineering, University of Hawaii, Manoa, Honolulu, HI 96822

[5] Mechanical Engineering, Iowa State University, Ames, IA 50011

[6] Environmental Engineering and Earth Sciences, Clemson University, Clemson, SC 29634

[7] School of Computer Engineering, Azad University, Science and Research Branch, Saveh, Iran

[8] Global Alliance for Rapid Diagnostics, Michigan State University, East Lansing, MI 48824

[9] Nano-Biosensors Lab, Michigan State University, East Lansing, MI 48824

*Corresponding author: Eric S. McLamore [emclamor@ufl.edu ☎ +1(352)294-6703]

**Corresponding co-author: Shoumen Palit Austin Datta [shoumen@mit.edu]

Abstract: We discuss the role of sensor analytics point solutions (SNAPS), a reduced complexity machine-assisted decision support (MADS) tool. We summarize the approaches used for mobile phone-based chemical/biological sensors, including general hardware and software requirements for signal transduction and acquisition. We introduce SNAPS, part of a platform approach to converge sensor data and analytics. The platform consists of a portfolio of modular tools which may lend itself to dynamic composability by enabling context-specific selection of relevant units, resulting in user-guided, context-specific, case-based working modules. SNAPS is an element of this variant configuration mechanism, where, depending on need, data analytics, statistical characterization and algorithms may be delivered to the data via embedded systems in devices, or sourced, in near real-time, from mist, fog or cloud computing resources. Convergence of the physical systems with the cyber components paves the path for SNAPS to progress to higher levels of artificial reasoning tools (ART) and emerge as data-informed decision support as a service (DIDA'S). Proof of concept examples of SNAPS are demonstrated for quantitative and qualitative data, each operated using a mobile device (smartphone/tablet) for data acquisition and analysis. We discuss the challenges and opportunities for SNAPS, centered around the value to end-users as a key performance indicator (KPI), for generic machine-assisted tools (MAT).

Keywords: sensor; smart systems; data analytics; cyber-physical systems; artificial reasoning tools; ART; drag and drop analytics; DADA; sensor-analytics packages; SNAP; machine-assisted tools; MAT; machine-assisted platform; MAP; knowledge graphs; trans-disciplinary convergence; machine-assisted decision support; MADS

This **DRAFT** copy of SNAPS is the 1st paper in a series of papers aiming to discuss *ART in SUGAR SNAPS PEAS for KIDS*
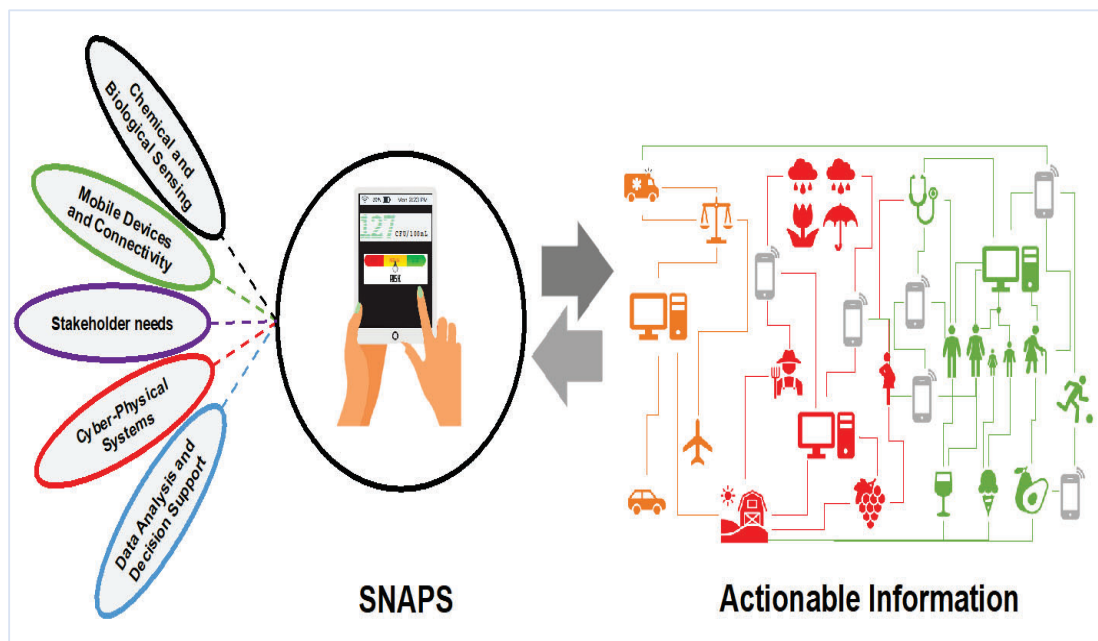
Figure 0. SeNsor-Analytics Point Solutions (SNAPS) represents a confluence of ideas, including chemical/biological sensing, mobile devices, and connectivity, cyber-physical systems, and data analysis for decision support. One intent is to deliver actionable information through contextually relevant applications using combinations of machine-assisted tools (MAT) and machine-assisted platforms (MAP).

1. Overview

We review select systems and sensor-analytics point solutions (SNAPS), which illustrates a confluence of ideas, including chemical/biological sensing, mobile devices, connectivity and cyber-physical systems (CPS), which may be combined with artificial reasoning tools (ART) and data-informed decision support (DIDA'S) systems. We provide an overview of signal transduction and engineering related to the recognition-transduction-acquisition triad, followed by a discussion of hardware and software elements for SNAPS. A summary of recent development of SNAPS may indicate short-term pragmatic opportunities sufficient to provide Pareto solutions (effective for 80% of the cases). We close by pointing out a few of the challenges, including the importance of making sense of data and how to deliver information on demand from data to end-users, before the quality of service (QoS) perishes, in the context of actionable information which may possess transactional value.

66 **2. Sensor Engineering**

67 *Chemosensing, biosensing, nanosensing may converge with partial autonomy*

68   A plethora of literature reviews describe the historical context [1], recent advances [2][3][4],
69 and futuristic ideas [5][6][7] related to development and application of chemosensors, biosensors,
70 and nanosensors [8][9]. These diagnostic tools have important applications across medical,
71 agricultural, and environmental domains. The sensor working mechanism established by the
72 International Union of Pure and Applied Chemistry in the 1990's serves as the design template.
73 The receptor coating on the sensor selectively binds target, a transduction event produces
74 measurable signal, the signal is acquired using specialty equipment. This sensing process, based on
75 the RTA triad, has been enhanced through the use of nanomaterials that improve detection limit,
76 speed and/or reversibility [10][11]. In addition, biomaterials are used to improve selectivity,
77 bandwidth, or facilitate actuation [12][13][14].

78

79   The primary application space for sensors/biosensors has been the analysis of unique targets
80 using relatively low cost, rapid detection platforms [15], including small molecules, viruses and
81 cells (amongst other targets). Recent works have focused on enhancing the mobility of sensors for
82 rapid on-site applications [16] by limiting the requirement for equipment or *post hoc* methods that
83 depend on formal labs. Most portable/handheld sensor efforts do not view chemo/bio/nanosensors
84 as a direct competitor for standard analytical laboratory diagnostics, but rather as a parallel tool to
85 trigger new questions or additional sampling to improve resolution and preserve data quality.
86 Attempting to use a handheld sensor to produce the accuracy and precision that is commonplace
87 in laboratory-based analytical techniques is in most cases a fool's errand, and a cost prohibitive
88 over-promise. What is realistic, on the other hand, is the development of low cost, light weight,
89 rapid diagnostic tools that can provide point solutions to match the specific context of the question
90 (the "low hanging fruit" from the tree of complex problems). Mobility of customized/personalized
91 sensors in an open-access format may prove to demystify the complexity of certain intractable
92 problems, raising the tide of knowledge, while providing service to communities in need and
93 enabling science to serve society (the masses, at large).

94

95   Mobile phone-based data acquisition systems are primary catalysts for mobility [17]. Smart
96 phone systems are available for optical transduction techniques such as fluorescence [18] and
97 surface plasmon resonance [19], in addition to electrochemical transduction techniques such as
98 voltammetry [20] and impedance spectroscopy [21]. While analytical capabilities have grown
99 exponentially in the last decade due to the rapid diffusion of tools such as machine learning
100 [22][23][24], there are only a few examples of mobile phone-based data analysis tools in the
101 literature [25] as most data analysis occurs on remote computers and not on mobile devices at the
102 point of contact. To maintain the integrity of end-user needs and ensure quality of service, mobile
103 phone-based sensors may be connected to remote analytics. Here, we introduce SeNsor-Analytics
104 Point Solutions (SNAPS), part of a platform approach for transforming sensor data into actionable
105 information using the mobile phone for data acquisition and performing near real-time, on-site,
106 edge analytics (**Fig 1**).
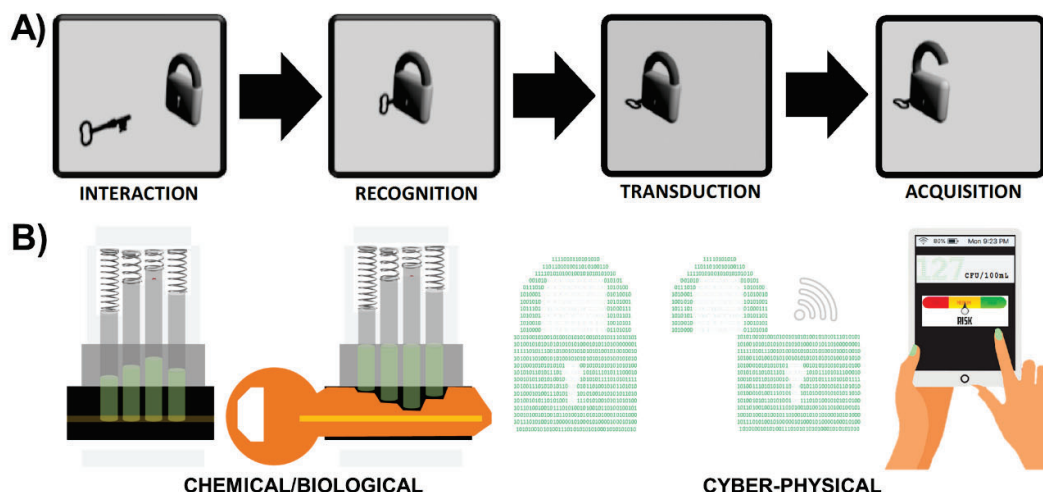
Figure 1. Sensor-Analytics Point Solutions (SNAPS) optimizes synergistic systems integration and connectivity between chemical/biological sensing with cyber-physical systems. **A)** Classical "lock and key" metaphor for sensor/biosensor part of the physical system in CPS. **B)** Sensor signal transduction (physical/chemical/biological component) and transmission to a mobile device coupled with in-network processing and on-site edge analytics (cyber component in CPS).

The RTA triad is governed by material choice

The affinity of the receptor for the target is the limiting factor for sensor function, and the importance of this first step in the RTA triad cannot be over-emphasized. Given that binding affinity and selectivity are the architects of the RTA triad, transduction is the platform for innovation. Intuitively, material choice dictates classification of device as either a sensor (use of abiotic materials), biosensor (biological or biomimetic materials), nanosensor (nanomaterials), or nanobiosensor (hybrid nano/biomaterials).

In addition to dictating these commonplace definitions, sensor material choice dictates critical factors such as performance, cost and quality of service (QoS). In its basic form, transduction is defined as a change in energy state, and there are two major classes of transduction that generates quantitative versus qualitative data, namely inherent and engineered transduction, respectively.

Engineered transduction involves specific binding between target and receptor but does not generate a measurable product without an extra "engineered" step. The thermodynamics of the system do not lead to favorable production of an active compound which can be directly quantified using acquisition equipment. An extra step is required such as addition of exogenous reagent(s) such as a label, or immobilization of a supporting material in close proximity to the recognition structure that results in a cascade which generates a measurable by-product. The most common example of engineered transduction is a lateral flow assay involving at least one binding protein (such as an antibody) (**Fig 2A**). This recognition structure is co-immobilized with a secondary structure that, upon binding of the target, undergoes a specific reaction and leads to a visible color

136   change [26]. In this type of transduction, the bonds between the target and recognition structure
137   are typically intact after the signal is acquired, leading to a significant amount of hysteresis (i.e.,
138   covalent antibody-epitope bonds). Due to the hysteric binding between target and receptor,
139   devices based on engineered transduction are typically not reusable as attempts to recover the
140   native binding chemistry of the receptor are uncommon. However, the latter may change and re-
141   usability may gain prominence if we can better appreciate the potential for engineering allosteric
142   triggers in the recognition mechanism [27][28]. At present, semi-quantitative data can be obtained
143   using engineered transduction approaches, if a sensor array is developed. The hysteric molecular
144   interactions restrict the data from being truly quantitative. Beyond color change, there are a host of
145   other transduction mechanisms that are not addressed in this paper (electrochemical, magnetic,
146   plasmonic, etc).
147
148   Sensors which autonomously produce quantitative data are classified as inherent
149   transduction. For example, an enzyme-ligand interaction is shown in **Fig 2B**. In this type of
150   transduction, binding of the target by the receptor leads to the production of a measurable by-
151   product with little or no hysteresis. No additional engineering is needed to obtain useful signal
152   correlated to the binding event, as the thermodynamics of the system indicate that the presence of
153   the target alone is the rate limiting step for energy state change. The formation of the product can
154   be directly correlated to the presence of a specific amount of target, with the causality of product
155   formation well described by system stoichiometry and thermodynamics. The most common
156   example of inherent transduction is the glucose biosensor for blood analysis. Glucose and oxygen,
157   present in blood, serve as activators of the enzyme catalyzed oxidation due to the enzyme GOx,
158   glucose-1-oxidase (beta-D-glucose:oxygen-1- oxidoreductase, EC 1.1.3.4). In this reaction, the
159   oxidation of glucose on the sensor surface results in the production of electrons, which are
160   measured using oxidative amperometry [29]. One major advantage of inherent transduction is that
161   the bonds between the target and receptor are typically destabilized during the transduction,
162   leading to diffusion of reaction by-products away from the binding site after formation. Reducing
163   sensor hysteresis facilitates development of reusable sensor chemistry, allowing continuous or in
164   line sensing.
165
166   Whether the recognition involves a biomaterial, abiotic, or nanophase material, in most cases
167   multiple chemical bonds occur between the target and the receptor material, and the strength of
168   these bonds governs the specificity, limit of detection (LoD), response time, and hysteresis of the
169   sensor. Mismatch of between material choice and intended application results in loss of quality of
170   service, and in some cases a complete lack of technology acceptance. Assays and sophisticated *post*
171   *hoc* analysis techniques can resolve some of this mismatch, within limits. To preserve and elevate
172   the quality of the outcome, selection of appropriate material(s) should be paired with optimized
173   sensing protocols and analytical techniques, discussed in the following section.
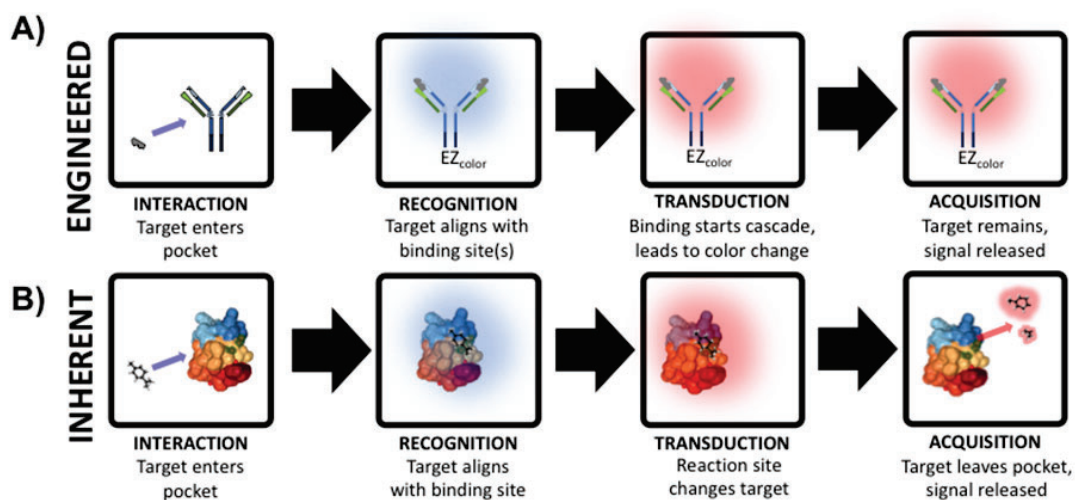
Figure 2. Development of chemical/biological sensor systems is based on either A)
Engineered transduction, where a cascade of reactions must be developed for obtaining
the signal that is specific to the interaction/recognition step, or B) Inherent transduction,
where activation energy is supplied by target and ambient environment. Structure of
protein in panel B courtesy of Mosbah et al [30].

3. Sensor-Analytics Point Solutions (SNAPS)

Overview

SNAPS consist of a biological/chemical/physical sensor directly interfaced with an analytical
tool. The green box in **Fig 3A** summarizes the RTA triad and displays a choice between the two
types of transduction discussed in the previous section. Once a receptor material is selected and the
appropriate transduction scheme is engineered, the process is coupled with acquisition equipment
to obtain signal (data). The blue box in **Fig 3B** shows the *post hoc* data analysis phase of SNAPS,
which aims to extract actionable information from sensor data. Contrary to the standard used in
sensor design, the analysis phase is less standardized, primarily due to lack of platform(s) for data
diagnostics, data quality, context, problem space, and query semantics [31]. As an example of a
common framework, Marr's framework is shown, which is a learning principle grounded in
Bayesian inference. Marr's analysis process flow has three interconnected steps: a computational
stage, an algorithmic or heuristic step, and an implementation step [32]. Analogous to the two
types of transduction previously discussed, the choice of a heuristic or algorithmic approach
should be directly linked to the problem context for maintaining quality of service. The
implementation step deconvolutes processed data using a relevance filter for producing
information.

In advanced SNAPS, the analysis phase may have an optional feedback control loop with the
sensor transduction step, which may be referred to as Smart SNAPS. For example, the temperature,
pH, electrical potential, or light intensity can be modulated to influence the sensor transduction

201 step based on information obtained from the data analysis phase. Active control of any phase in
202 the RTA triad qualifies as a Smart SNAPS, but interfacing with the transduction step is the most
203 logical route for adding value.
204
205   In SNAPS, acquisition and analytical processing occurs at the edge by deploying a mobile
206 platform of tools using a smartphone or a tablet, or other similar devices as mobile hosts. The next
207 section demonstrates examples of these hardware and software tools in the current literature.
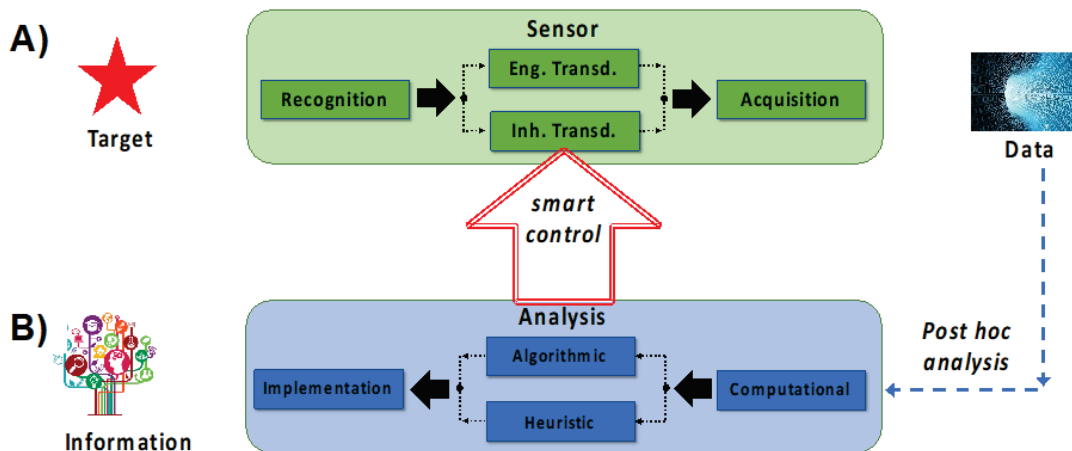
208

209 **Figure 3**. SNAPS attempts to transform data into information based on convergence of
210   two distinct areas, namely sensing and analytics. The framework for these two areas is
211   described by: **A)** standard sensor development guided by RTA logic, and **B)** data analysis
212   using the usual tools (computational, algorithmic, statistical). Smart control can be
213   achieved when data from the analysis step actively controls (auto-actuates) at least one
214   process within the RTA sensor triad.

215 *SNAPS Hardware and Software*

216   **Fig 4** shows an example of the hardware and materials used for the development of SNAPS.
217 The example of engineered transduction (top of **Fig 4**) demonstrates engineered transduction for
218 diagnosis of tuberculosis (TB) via detection of acid-fast bacilli in sputum samples. Biorecognition
219 is grounded in principles of glycobiology, where target cells are labeled by glycan-coated magnetic
220 nanoparticles (GMNP) [33][34]. In this example, a neodymium magnet is used to separate the
221 particle-cell aggregates to facilitate rapid determination of acid-fastness and cording properties of
222 captured mycobacteria. The TB test also employs Gram staining to provide visual confirmation.
223 Smartphone based optical systems [35] or complex microfluidic systems [36] may be used for on-
224 site image analysis, and image processing algorithms [37][38][39][40], may be used for improving
225 accuracy and providing decision support.
226
227   In another example (bottom of **Fig 4**), inherent transduction is demonstrated for approaches
228 such as the graphene oxide-biogenic amine nanobiosensor recently developed by Vanegas et al
229 [41]. In this example, an enzymatic biosensor was developed based on diamine oxidase, which was
230 tethered to a laser scribed graphene (LSG) electrode decorated with nano-copper. Upon

231 recognition of the target ligand within the enzyme binding pocket, oxidation is carried out to
232 produce hydrogen peroxide as a by-product. The peroxide is deprotonated under an operating
233 potential of +500mV to produce electrons, measured using oxidative amperometry. Inherent
234 transduction uses materials such as oxidase which contain both active binding site(s) for
235 recognition and catalytic reaction site(s) for transduction, with the pockets remaining intact after
236 reaction (limiting hysteresis). Signal acquisition utilizes a handheld potentiostat connected to a
237 mobile phone such as the ABE-STAT tool [21]. The support vector machine (SVM) learning
238 (SVML) classification system [25] may be also applied using the same mobile phone via the Jupyter
239 notebook open source portfolio of machine learning tools. There are a myriad of other approaches
240 for optical smartphone sensing [19][42][43] as well as electrochemical sensing [20][44].
241
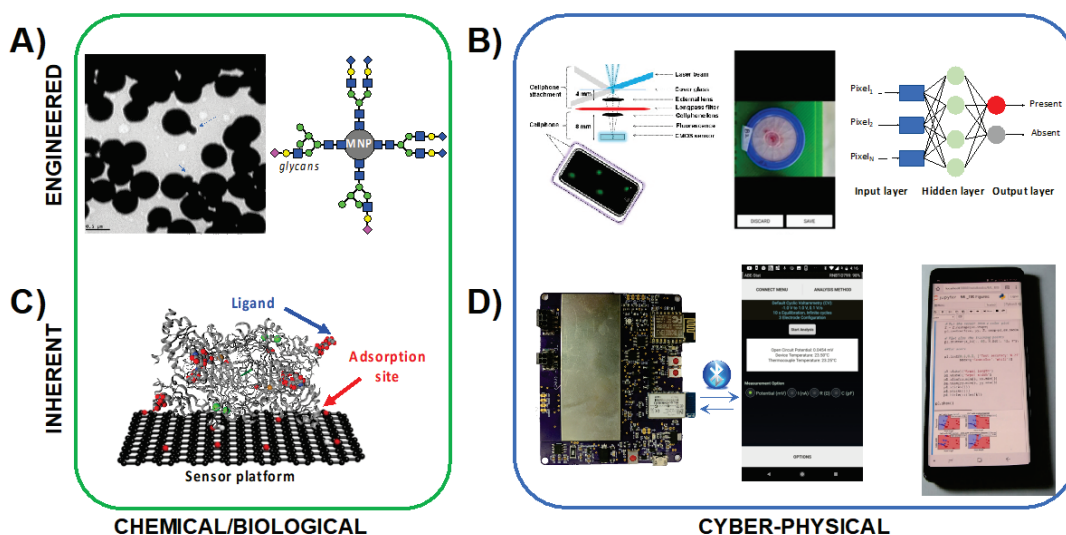
242

243 **Figure 4.** Examples of hardware/materials that may be used for SNAPS for engineered
244 transduction (top) and inherent transduction (bottom). These select examples
245 demonstrate the convergence of chemical/biological sensors with cyber-physical systems
246 for delivering computational capabilities and *algorithms to sensor data* (rather than
247 *sending* the data for analytics, in this paradigm the analytical tools are *sent* to the data at
248 the edge, on a mobile device). **A)** Bacteria-specific magnetic nanoparticles provide dual
249 functions of labeling and enhancing aggregation in detection of target cells. **B)** Cyber-
250 physical tools for engineered transduction may include smartphone-based microscopes
251 for cell imaging, the images are processed by embedded algorithms and the outcome
252 "informs" the decision support system. **C)** Oxidase-based biosensors enable inherent
253 transduction, limiting hysteresis and enabling reusability. **D)** Handheld potentiostats
254 may be used to acquire electrochemical sensor signals, and embedded or cloud-based
255 remote analytics such as SVML [25] may be used to provide analysis and decision
256 support. Photograph of glycan-functionalized magnetic nanoparticles courtesy of Bhusal
257 et al [33]. Crystal structure of diamine oxidase courtesy of McGrath et al [45].
258 Fluorescent smartphone sensor platform courtesy of Wei et al [35]. Image of colorimetric
259 E. coli test and machine learning knowledge graph from Gunda et al [40].

This **DRAFT** copy of SNAPS is the 1st paper in a series of papers aiming to discuss *ART in SUGAR SNAPS PEAS for KIDS*

260   It is obvious that sensor data, and thus the hardware to collect the data, are core competencies
261 required to fuel SNAPS or any equivalent tool in the portfolio of machine-assisted tools (MAT).
262 Without data, subsequent progress from SNAPS to higher order decision support may be
263 impossible. A wide range of commercial and custom software are available for cloud-based
264 analytics [46][47][48]. Drag and drop analytics [49][50][51][52][53][54] is one such tool of
265 growing interest. One example of a drag and drop analytic software is the tool developed for
266 quantifying uncertainty in data exploration, or QUDE. QUDE automatically quantifies the
267 different types of uncertainty/errors within data exploration pipelines [55]. The automation feature
268 in this tool is based on the following workflow: data extraction, data integration, data processing,
269 exploratory queries, machine learning, and finally interpretation. QUDE is not intended to
270 represent a solution agnostic of all problems, but rather demonstrates an approach that may serve
271 as a starting point to connect chemical/biological sensors to analytics in real time. Evolution of
272 data analysis from extraction to visualization, even in a drag and drop *modus operandi*, requires a
273 deep understanding of the context and is highly problem specific. Visualization tools (such as the
274 volcano plots in **Fig 5**) are information-rich presentations of complex datasets which may facilitate
275 use of a tool in multiple application domains, but the information must be comprehensible to
276 users/stakeholders to provide any actionable information.
277
278   Knowledge graph algorithms, in combination with statistical analysis and machine learning
279 (for example, feature engineering, extraction and selection) [56][57], are elements likely to
280 improve the outcome from SNAPS and enable the gradual evolution of SNAPS to become a higher
281 order tool to facilitate data-informed decision as a service (DIDA'S), as discussed in the following
282 section.
283
284   The plethora of tools reviewed here and elsewhere may operate in harmony in specialized
285 facilities (such as academic research centers). However, the real value of the convergence is at the
286 hands of end-users who may lack specialized knowledge of software or systems. Without lucidity
287 as a guiding light in the design process, any meaningful application of these tools may remain an
288 utopian expectation. Adoption largely depends on creating user interfaces no more complicated
289 than a menu of choices, for example, the type of variant configuration that enables a customer to
290 customize a laptop. Thus, the paradigm of "plug and play" must be at the front and center of this
291 discussion in order to "hide" the complexity behind simple drag and drop features which may
292 empower end users to efficiently interact with these tools [58][59]. Simplified drag and drop tools
293 for SNAPS may exponentially accelerate global demand for these tools. Democratization of access
294 through Lego-esque modular drag and drop interfaces [60], may pave the way for "drag and drop
295 analytics" (DADA) for mobile decision support systems and partial autonomy. DADA and SNAPS
296 have enormous application potential, not only in the agro-ecosystem but in any domain, for
297 example, healthcare, manufacturing [61]**,** finance, utilities, logistics, transportation and retail.
298
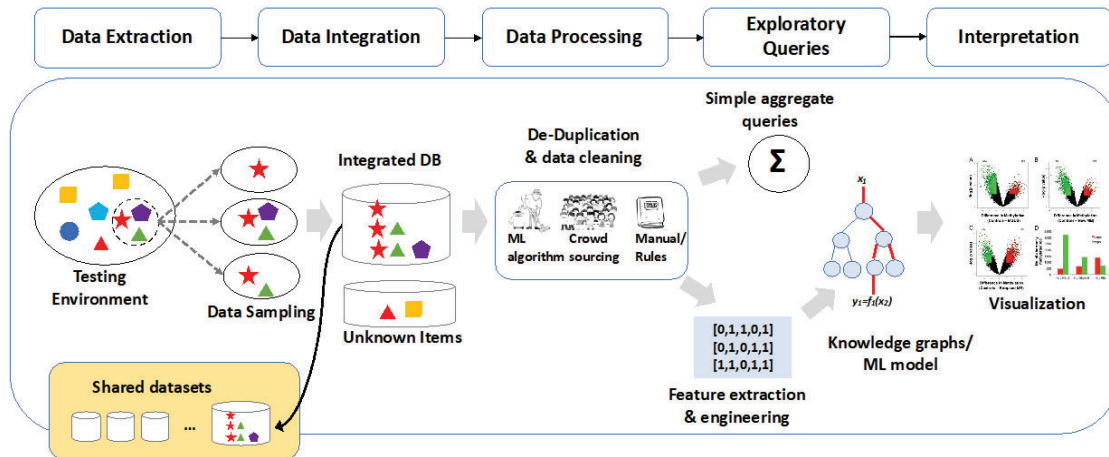
Figure 5. Software tools such as drag and drop analytics (DADA) may enable cloud-based analytics for SNAPS. The tool shown here automatically quantifies the different types of uncertainty/errors within data exploration pipelines (image from Chung et al [55] modified to match context of SNAPS). Diagram shows workflow (top) and an example of a pipeline (bottom).

4. Auto-actuation and partial levels of autonomy for low-risk automation (SARA Paradigm)

Autonomy is a framework that emerged from intelligent control and systems theory dating back at least a half century [62]. The specific technology need predicates the architecture of the system (including both hardware and software), and not all sensors or sensor systems are required to be involved in higher levels of autonomy [63][64]. For example, the purpose, architectural details, system functions, and characteristics for unmanned *terrestrial* vehicles are different compared to unmanned *space* vehicles [65]. One unifying attribute of these systems is the need to sense (sensors on the front end of the process) and then analyze (analytics on the back end) before the system can respond, i.e., SNAPS equivalent in the sense-analyze-response systems (SARS) paradigm. Lessons from higher level autonomous systems may inform how SNAPS may be better engineered and optimized to deliver value (albeit, in a very different context, when compared to the specifications for levels of autonomy in automobiles or the aerospace industry).

In the context of SNAPS, the traditional six levels of autonomy (*see supplemental Figure S1*), may inform design. In the lowest level of autonomy (simple), human interaction is required for direct control of the sensor system(s) and/or manual off-loading of data for *post hoc* analysis. For example, deployed buoy systems are common in environmental studies of aquatic chemistry [66], which represents the current state for most sensor data, which is simply a report of raw data. The latter may be referred to as simple user-guided activity report or the SUGAR paradigm.

In the 2nd level of autonomy (assisted), a high degree of human interaction is required, but at least one aspect of SNAPS (either sensing or analytics) is capable of performing task(s) without *de novo* synthesis of a map. These tasks may achieve prescribed objective(s), adapt to environmental changes, or develop new objectives. For example, Rong et al [25] recently developed an open

329   source mobile-phone based analytics protocol for analyzing impedance data acquired from a
330   nanobiosensor. The primary objective of the tool is to perform the first layer of analysis in
331   development of a SVML classifier to analyze impedance data (in lieu of equivalent circuit analysis).
332   The mobile phone-based tool automates selection of classifier type using principal components
333   analysis, and subsequently automates selection of hyperplane parameters (optimizes the support
334   vector classifier and support vector regressor functions across the selected hyperplane). While this
335   MAT does not perform decision support or provide validation layer(s), it may qualify as machine-
336   assisted automation, particularly if the impedance data is acquired using the same hardware and
337   the sensing/analysis processes are linked for on-site edge analytics (that is, analysis, at the edge).
338

339   The classical 3rd level of automation, *partial autonomy*, may be achieved through remote
340   control of SNAPS related features, including sensing, data download, and some form of data
341   analytics such as heuristic risk assessment. The outcome may trigger a low-risk set of logic tools to
342   execute workflow which sets into motion an auto-actuation function. By embracing and
343   accomplishing auto-actuation, the concept of SNAPS evolves to address the SARA (sense, analyze,
344   respond, actuate) paradigm. For example, SNAPS may auto-adjust water flow rate in irrigation
345   pumps (by temporarily overriding a pre-set routine flow rate) based on updated moisture data
346   (from field sensor) and refreshed external weather data predicting rapidly advancing thunderstorm
347   with imminent rainfall. Hence, control systems like SNAPS, manifesting elements of the SARA
348   paradigm are derived from principles of bio-mimicry. Feedback control (activation/inhibition) is
349   the bed-rock of biological systems in maintaining homeostasis, cellular equilibrium and
350   harmonization of physiological balance [67].
351

352   Higher levels of automation are beyond the scope of this discussion. For partial automation,
353   SNAPS shall increasingly rely on integration of logic structures, for example, integrating output
354   from one or more artificial reasoning tools (ART) to support auto-actuation (SARA). Integration
355   of logic and ART in the SNAPS architecture indicates a departure from simple point solutions
356   (SNAPS) and an upstream march of ART, ushering new layers of convergence beyond SNAPS. In
357   this description, we focus on two major categories: i) sensors with engineered transduction
358   coupled with heuristic analysis of qualitative data (SNAPS-ART), and ii) sensors with inherent
359   transduction coupled with algorithmic analysis of quantitative data (referred to as SNAPS-
360   DIDA'S), summarized briefly, in the following section.

361   **5. Design of SNAPS for Decision Support Systems**

362   If auto-actuation or partial automation is expected as an outcome, then, according to the
363   SARA paradigm, it is necessary to integrate elementary logic layers, relevant to performance/event,
364   to enable SNAPS to execute actions by using/combining contextually relevant outputs from ART.
365   The dynamic nature of ART elements may be sourced from a holding platform or repertoire for *ad*
366   *hoc* composable machine-assisted tools (MAT), which applies basic pseudocode, with simple logic,
367   to provide case-specific outputs, sufficient to execute low risk actions, which are (highly) fault
368   tolerant. To provide guidance in development of future SNAPS, we may organize tools into two
369   categories that are designed to meet user needs while maintaining an appreciable quality of service

370   (QoS). The first category utilizes qualitative sensors based on engineered transduction coupled
371   together with heuristic analysis to produce a version of ART (**Fig 6A**). SNAPS-ART, is designed to
372   provide near real-time management suggestions, such as using (single) sensor data to determine
373   whether a particular sample is above or below a threshold value pre-set by a regulatory agency. The
374   assumption of high fault tolerance and low risk are pivotal to develop and deploy SNAPS-ART.
375

376   To optimize QoS versus development costs, acquisition of qualitative data for SNAPS-ART
377   uses engineered transduction techniques and heuristic classification to satisfy user expectations
378   with a binary output (rapid YES/NO test). In terms of active control features, SNAPS-ART is
379   rudimentary, with a few discrete/distinct actions (turn on/off a subsystem) determined by simple
380   non-overlapping binary outputs based on input from SNAPS. SNAPS-ART is not intended to be a
381   comprehensive diagnostic tool, but designed for triage or rapid screening, where additional testing
382   is required to confirm/validate results. It is possible to use ART for semi-quantitative purposes that
383   depends on other combinatorial factors (for example, controlling rate of flow), but within reason.
384

385   The 2nd category of SNAPS (**Fig 6B**) utilizes quantitative sensors based on inherent
386   transduction coupled with algorithmic analysis, collectively referred to as SNAPS-DIDA'S.
387   Contrary to SNAPS-ART, this category is designed for decision support under the assumption of
388   low fault tolerance and moderate risk. Rather than instantaneous results classified by heuristic data
389   analysis, SNAPS-DIDA'S involves a dynamic and/or reiterative analysis of streaming data from
390   sensors and feedback logic interfaces for processed data, for example, active control features using
391   a case-specific subset of tools from a super-set of MATs and MAPs (machine-assisted tools and
392   machine-assisted platforms). Optimization (from menu of choices and range of values) for each
393   variable requires computational rigor and resources to extract *context-specific* variant
394   configurations expected of SNAPS-DIDA'S. On the other hand, workflow middleware as the logic
395   application or control layer may suffice for SNAPS-ART. Sophisticated decision support software
396   with decision trees executing embedded logic is one option which may be user-directed [68][69].
397   The latter may be enabled by a drag and drop assembly from a portfolio of modular tools under the
398   umbrella of MAT and MAP. Another option is to present these choices to a human-in-the-loop
399   who may exercise some form of exclusion/reduction to narrow the search space (number of
400   choices) from the MAT/MAP menu based on experience and knowledge [70][71]. The preferred
401   option is the development of a parallel agent-based system (ABS) which may be part of a multi-
402   agent system (MAS) [72][73]. The agent is expected to specific for certain pre-determined
403   functions and able to replicate/reason a few key choices (select functions), in a manner
404   approximating the human-in-the-loop. ABS may be void of the human ability to handle exception
405   management. The range of options for the Agent may be restricted by its arsenal of information
406   and logic rules, due to limits of learning and training software Agents, especially about how and
407   when to use which tool. Because of the cognitive boundaries of deterministic design, it will be an
408   egregious error to expect any training tool or machine learning routine to educate an Agent to
409   deliver support in non-deterministic scenarios. The latter makes it mandatory to recognize
410   boundaries of "artificial" systems and maintain provisions for humans-in-the-loop, by design, for
411   non-deterministic instances.

412   In the context of SNAPS-DIDAS, agent-based systems begin to function upon receiving input
413   from SNAPS. Using logic capabilities (learned, trained), agent(s) determine which tool, or sets of
414   tools, may be necessary to execute the action or partial automation SNAPS expects to trigger.
415   Agents are limited by the tools contained within MAT and MAP, unless embedded logic provides
416   the option to place a remote function call (RFC) to a cloud repository to source other tools or
417   algorithms from different systems. Including and enabling this feature is vital for agent(s) to
418   "discover" contextually relevant elements suitable for the case at hand (perhaps enabling a
419   potential path to higher levels of automation). Unrestricted discovery demands interoperability
420   between systems to source tools and functions in a manner that is location (server) agnostic.
421   REST (representational state transfer) is a lightweight architectural alternative to mechanisms like
422   RPC (Remote Procedure Calls) which helps to connect between distributed hypermedia systems.
423   RESTful web services (conforms to REST architecture) facilitate interoperability between
424   computer systems and RESTful APIs (web service APIs that adhere to the REST architectural
425   constraints) are quintessential for "plumbing" the network to catalyze connectivity [74].
426

427   Knowledge discovery/search functions of machine-assisted systems (MAS) are key
428   performance indicators (KPI) which are inextricably linked with QoS. Synergistic integration with
429   external tools and modules calls for interoperability between platforms, which are influenced by
430   standards and architecture. Hence, automation of SNAPS-DIDA'S is far from trivial, and we are
431   only beginning to scratch the surface in terms of the confluence of ideas necessary to transform
432   this vision into reality.
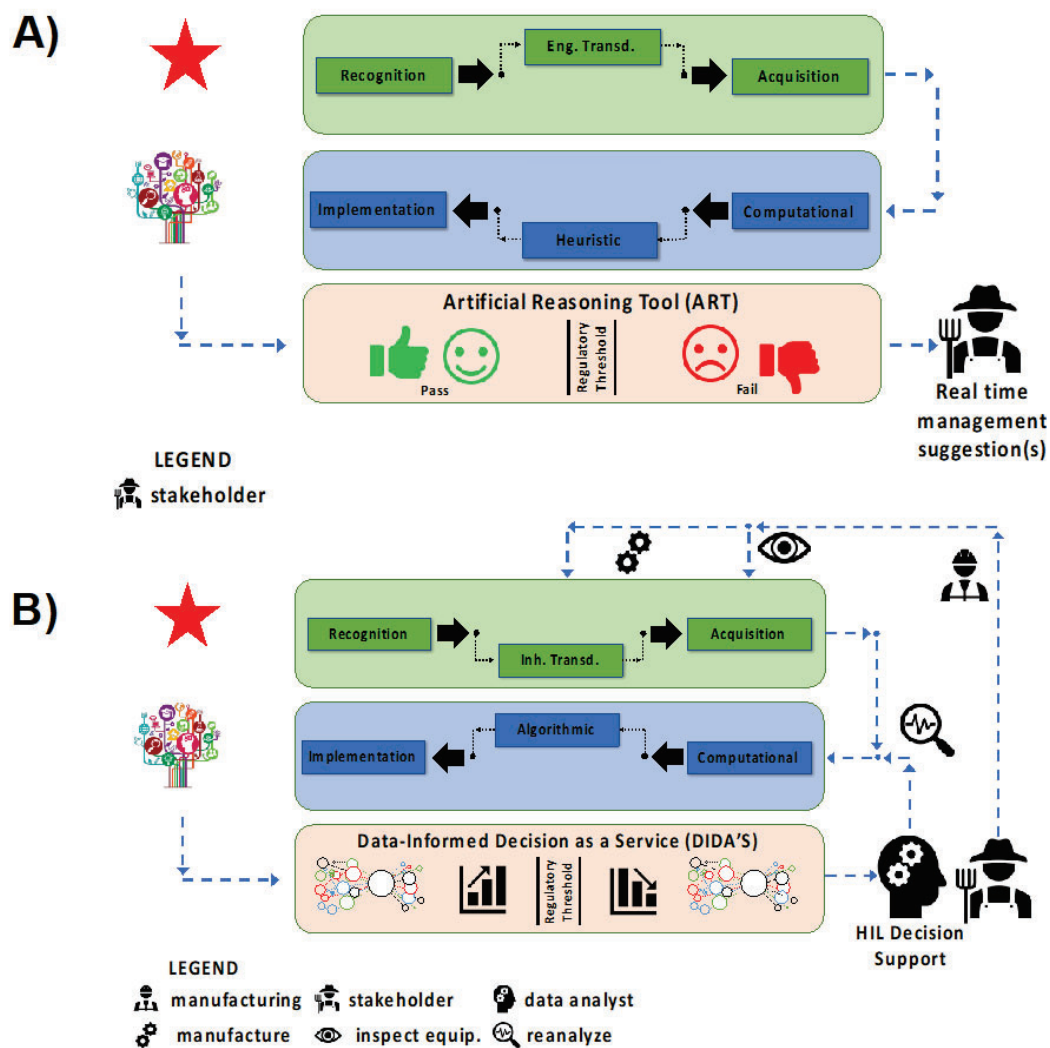
Figure 6. Classification of SNAPS based on end-user expectations. **A)** SNAPS-ART produces qualitative, or semi-quantitative sensor data based on engineered transduction for analysis using heuristic analysis tools to provide management suggestions. **B)** SNAPS-DIDA'S produces quantitative, streaming data for algorithmic analysis to be implemented in a variety of decision support paradigms, which may include human-in-the loop and/or agent-based systems. The diagram provides a map for avoiding mismatch (i.e., loss of quality of service) between sensor chemistry and application needs but are not intended to be formulaic or dogmatic. Being dynamic and agile is essential.

The theoretical boundary between ART and DIDA'S is blurry, at best. The classification of the two systems into discrete boxes (**Fig 6**) is not intended to be reductionist. Perhaps it may trigger ideas about future improvements and innovations. The distinction between ART and DIDA'S may be made in terms of the data that must converge or the degree to which data fusion may be necessary when rendering the decision or recommendation. ART may assist a rapid-response system which aims to solve low risk problems, with a few data sources and data dependencies,

448 using either qualitative or quantitative sensors, matched with heuristic analysis. For example,
449 *qualitative* SNAPS-ART may provide the instruction to turn off the irrigation water pump if [a]
450 the rate of change of 80% of the soil moisture sensor readings fall above/below a given range of
451 values or [b] if the data from sensor(s) indicates that the rainfall rate is above a certain value. For
452 *quantitative* SNAPS-ART, the instruction may be to monitor and turn up/down the rate of
453 irrigation water flow, by grids/zones, depending on the soil moisture sensor readings, if the sensor
454 data falls above/below a range of values. The tool may refer to the logic instructions in a lookup
455 table which recommends water flow rates which are recommended (appropriate context) for given
456 soil moisture content and also integrates weather data.
457
458   DIDA'S may be viewed as a mutiny of multiple ART units, each vying to contribute its data.
459 At each gateway or node in the DIDA'S platform, there are Agents which are queueing, to be
460 triggered by specific data strand/stream, to start its search and discovery process, to identify what
461 tools must be used, which other databases or data resources must be accessed, in order to satisfy
462 the context of dependencies and when/how to feed the results from the search and discovery to a
463 higher level Agent. This situation is analogous to dynamic composability of tools, triggered by
464 data, in a manner similar to application dependent networking (ADN), the underlying principle
465 necessary to connect two mobile phone users in diverse environments [75].

466 **6. Proof of concept SNAPS-ART**

467   **Fig 7** demonstrates an example of SNAPS-ART for heavy metal analysis coupled with risk
468 assessment. In this design, a sensor with engineered transduction (qualitative data) is coupled with
469 a heuristic risk analysis tool (hazard quotient [HQ] indicator) for monitoring and assessing risk of
470 mercury exposure in drinking water in communities located near small scale artisanal gold mining
471 regions [76]. In this example, a nanosensor was developed based on LSG electrodes decorated with
472 anchored nanocopper for measuring ionic mercury (Hg^{2+}) via stripping voltammetry [77] (**Fig
473 7A**). Rapid screening of water samples for mercury contamination is highly useful, but the value of
474 sensor data is inconsequential without information on how compounded factors, such as body
475 weight, ingestion rate, and length of exposure contribute to overall risk. A mobile app was
476 developed in R language using the heuristic hazard quotient (HQ) method used by global
477 regulatory agencies [78]. MIT App Inventor was used to create a smartphone app, using drag and
478 drop techniques and Blockly modular tool (see supplemental section for code and details regarding
479 app development). **Fig 6B** displays the graphic user interface and a sample output for the SNAPS-
480 ART tool, where users input drinking water ingestion rate, body weight, length of exposure, and
481 age. The app captures Hg^{2+} levels (ppm) obtained from the sensor, and uses the framework
482 established by the US Environmental Protection Agency and the World Bank to calculate a HQ
483 score [79][80][81]. Using the EPA standard HQ threshold [82], HQ >1.0 indicates higher risk of
484 potential adverse health effects increases, while a score <1.0 indicates low risk.
485
486   It is significant to grasp that the suggestion offered to the end-user is ***not*** the raw sensor data
487 output but a value (HQ) which ***meaningfully integrates*** other types of data to synthesize user-
488 specific ***information*** which the user may choose to use, effectively (or not). Thus, SNAPS-ART

This **DRAFT** copy of SNAPS is the 1st paper in a series of papers aiming to discuss *ART in SUGAR SNAPS PEAS for KIDS*

489   provides a real time **suggestion** to the user for health and wellness purposes as a result of
490   convergence of data including international standards for heavy metals contamination of drinking
491   water and healthcare statistics, relevant to the end-user and her environment. Detecting heavy
492   metal in water [83] and access to data was facilitated when text messages (SMS) with geo-code
493   (GPS location) were enabled to retrieve test data [84]. Improving the quality of raw data [85] is of
494   paramount importance but the end-users are in quest of answers (information) rather than
495   numbers (data). Simple and direct decision support using mobile devices is key for resource-
496   limited populations [86].
497
498   This example of SNAPS-ART highlights this quintessential **_transition_** from *data* for experts to
499   *information* for end-users. The HQ output significantly increases the **_value_** of the raw sensor data
500   by *combining* data in context of human-specific factors and micro-environment, to provide
501   actionable information relevant to precision public health. In addition, based on the HQ score, the
502   embedded artificial reasoning tool (ART) provides options to seek additional screening at a
503   verified laboratory, if the sample value triggers a set of logic/rules based on relationships between
504   data, perhaps revealed due to data association and/or data fusion. The latter is a critical feature
505   because the outcome has the potential to alert the user to recognize that the suggestion is not only
506   displaying a number but requesting secondary validation (visit to a primary care, formal lab tests).
507
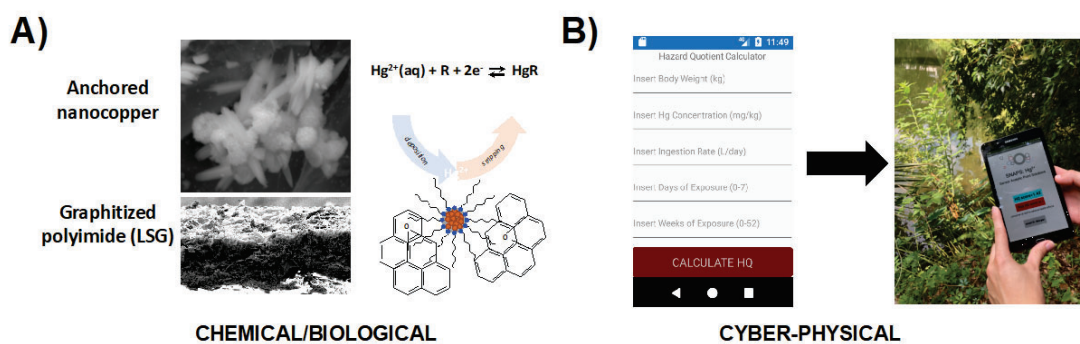
508
509   **Figure 7**. Proof of concept demonstration for SNAPS-ART in heavy metal analysis of
510   water. **A)** Hg^{2+}-selective nanosensor based on LSG and nanocopper developed by
511   Abdelbasir et al [77]. **B)** Photo of heuristic analysis tool for calculating risk of mercury
512   exposure (hazard quotient calculator).

513   **In** another application (Fig 8) the SNAPS-ART platform aggregates impedimetric sensors for
514   detection of pathogenic bacteria in food samples. The *Listeria monocytogenes* biosensor developed
515   by Hills et al [87] is used as a demonstration (**Fig 8A**), and an ART tool was developed using
516   machine learning (R programming language). The ART tool for *L. monocytogenes* detection is
517   grounded in binary classification using bagged random forest written in R, with the smartphone
518   app created using MIT App Inventor. The program (see supplemental section for code) reads a raw
519   impedance data file from the biosensor, converts the data to the necessary form for machine
520   learning classification, optimizes hyperparameter values, and then uses machine learning
521   techniques to compare the sample to a training library; other methods are feasible as described by

522   Rong et al [25]. The tool is used for predicting whether the food sample (vegetable broth in this

523   case) may be contaminated or is safe, and how an user may seek secondary validation, if sample is

524   positive (**Fig 8B**); see supplemental section for details and screenshots of app. The major benefit of

525   this tool is the avoidance of computationally expensive (and time intensive) analytical methods

526   such as equivalent circuit analysis. While some equivalent circuit models, such as the Randles-

527   Ershler circuit, provide some description of the physical meaning for each circuit element related

528   to an impedimetric biosensor, in most cases more complex models are used and parameters are

529   tuned with chi-squared (χ^2) fitting. If not used with expert guidance, equivalent circuit analysis

530   leads to significant errors in both interpretation and accuracy. It may be cost prohibitive, too. For

531   pathogens such as *L. monocytogenes*, the threshold for contamination is one live cell in a food

532   sample, and thus speed and accuracy of the tool are pivotal for rapid screening. Use of machine

533   learning (ML) tools to perform rapid screening, at the point of use, with a smartphone or tablet,

534   increases the **_value_** of the biosensor data and offers near-real time actionable *information* to the

535   end-user, agnostic of location or access to a laboratory test facility.
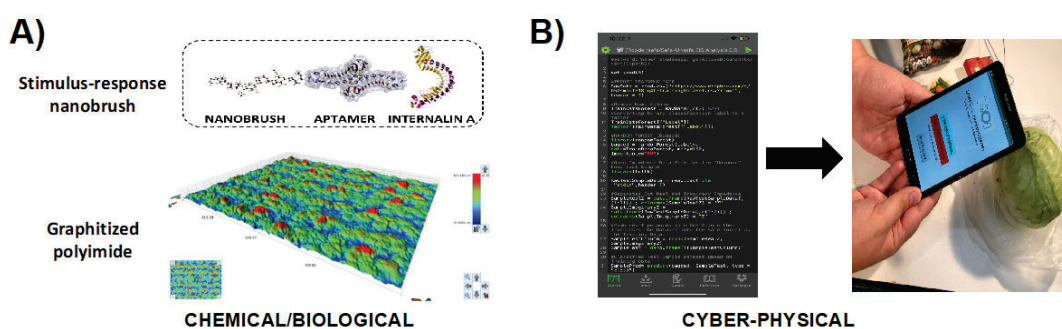
536

537   **Figure 8**. Proof of concept demonstration for SNAPS-ART in food safety analysis

538   (vegetable broth). **A)** *Listeria monocytogenes* biosensor developed with stimulus-response

539   polymers and DNA aptamers by Hills et al [70]. **B)** Screenshot and photograph of

540   machine learning analysis tool for determining whether sample is contaminated based on

541   an index score derived from machine learning analysis.

542   The benefits of the approach in **Fig 7-8** may be extended by adapting the sensor RTA scheme

543   to allow the tool to detect and alert users about other targets (eg other heavy metal contaminants,

544   such as lead, cadmium, and arsenic or other biomolecule targets, such as viruses). SNAPS-ART

545   may be used for detection and diagnostics for a plethora of contaminating agents not only in liquid

546   (as shown here), but in any other medium as long as the analyte is presented in a form that binds

547   with the sensor material. The value of this approach increases exponentially if combined with

548   mobility (smartphone). The latter enables sourcing sensor data for myriad of analytes from any

549   environment where humans (citizen science) or drones may reach or interact with the sample. In

550   our approach, we have eschewed the use of high cost sensors, to highlight the potential for

551   diffusion of low-cost tools to enable democratization of data and distribute the dividends from

552   decision support to serve community-specific needs. To acquire, curate, analyze and extract useful

553   information from sensor and other data, we advocate synergistic integration with MAT and MAP.

554   SNAPS is a tiny first step in that direction.

555 <div style="text-align:center">**Table 1**. Challenges and Opportunities for SNAPS</div>

Challenges	Opportunities
Extraction of information from sensor data for real time decision support	• Development of SNAPS-ART tools using established regulatory standards as a guide
Controlling or modulating sensor hysteresis *in situ*	• Integration of smart materials on sensor surface (e.g., stimulus-response polymers) • Rudimentary control over system performance through the use of sense-analyze-respond-actuate (SARA) systems
Mobility and connectivity in agricultural and environmental systems	• Improvement of low latency communications • Deploy low power [a] sensor systems and [b] data management systems
Integrating SNAPS with standard platform	• Establishment of data management systems based on lessons learned from other systems such as integrated clinical environment (ICE) • Establishment of standards in architecture
Development of data informed decision as a service (DIDA'S)	• Establishment of SNAPS-ART as a common tool • Integration of drag and drop analytics (DADA) and agent-based systems (ABS) • Dynamic/reiterative analysis of streaming data from sensors • Demonstration of feedback logic that interfaces with processed data • Dynamic composability of tools triggered by data (application-dependent-networking) • Database discovery or data resource discovery by agents using embedded logic (e.g., remote function call, RESTful APIs) • Use of logic capabilities (learned, trained) and agent(s) to determine which tool, or sets of tools, are required for analyzing any given problem

556
557

558   **7. The Path Forward?**

559   The pivotal role of sensors, data, and information in decision support and partial automation
560   or auto-actuation is of critical importance in any field, including biosystems engineering. SNAPS
561   represents a confluence of ideas and is the foundation for making sense of data by adding value to
562   sensor data. Basic sensor design choices, described in this review, dictate the value and quality of
563   service for SNAPS. The lowest common denominator is to match the type of sensor transduction
564   (engineered or inherent) with an analytical approach (heuristic or algorithmic) to meet the needs
565   of the end user with respect to a baseline quality of service (QoS). In a more complex example,
566   sensor engineering for controlling or modulating hysteresis is required, if users expect real time, *in*
567   *situ*, sensor data connectivity to data analytics. SNAPS can offer design choices and suggest rules to
568   adopt (embedded logic) key performance indicators (KPI) which may be necessary to guide what
569   constiutes actionable information based on raw sensor data. However, it is essential to remain
570   cognizant of the fact that not all data may contain information and not all information may be
571   actionable or possess transactional value. Hence, the quality of data determines the quality of
572   service (QoS outcome).

573

574   SNAPS and ART are elements of the MAT and MAP trends, in terms of the quest to deliver
575   value from data analytics (in this case, with heuristic and qualitative sensors). Rudimentary control
576   over system performance through the use of SARA-driven Smart SNAPS, to auto-actuate select
577   system components, may be an example of emerging applications for agro-ecosystem,
578   environmental health, as well as public health and healthcare, especially remote diagnostics and
579   preventative health.

580

581   Beyond the SNAPS-ART paradigm, we hope to reach DIDA'S but its maturity is questionable,
582   at best. As these tools mature, the true value may be realized through the interaction of agents (i.e.,
583   an agency) which embraces the collective optimization of performance (P) in the context of the
584   environment (E) where processes may be influenced by actuators (A) that depend on information
585   from sensor (S) data (thus, PEAS).

586

587   PEAS is mnemonic borrowed from the literature on ABS designed (modeled) to address
588   systems performance [88] that is focused on the convergence of percepts, environment, actuators,
589   and sensors (PEAS). The principle of PEAS are supposed to be the pillars on which we may build
590   "machines that work for us" versus "cogs" in the wheel - SNAPS, ART, DIDA'S - which are
591   examples of "tools with which we work" [89].

592

593   PEAS are goal-dependent strategic perspectives arising from the conceptual pot of alphabet
594   soup each with its array of dynamic push-pull elements and user-directed levers, which may be
595   used in any combination, to accomplish short term tasks (SNAPS) versus long term attempts to
596   orchestrate systems performance (PEAS). This interrelationship may be analogous to components
597   of the engine (SUGAR, SNAPS, ART, DIDA'S) which are essential and dependent for the function
598   and performance of the "whole" vehicle (PEAS).

599

600   Aggregating data and information for systems performance (PEAS) is the Holy Grail. In some
601 instances the "whole" picture is the *only* relevant picture. One attempt in biomedical engineering
602 is referred to as the integrated clinical environment (ICE) effort [90], which strives to harmonize
603 data interoperability between all sub-systems to focus on the "whole patient" rather than isolated
604 parts. These concepts (PEAS, ICE) may serve as beacons for innovation in agriculture, energy,
605 environment, or other verticals areas, where a *tapestry* of solutions may be more valuable than
606 *point* solutions. Isolated solutions in the medical systems may lead to errors due to medical device
607 interoperability issues. The latter is often fatal, claiming as many as 250,000 lives per year, in the
608 US alone, and is the third leading cause of death in the US [91].
609

610   While no specific decision support framework is shown here, SNAPS is the first step toward
611 DIDA'S, which may rely on tools such as drag and drop analytics (DADA) and models using ABS.
612 One of the underlying themes in all cases is the use of mobility and low latency signal transmission
613 (for example, future potential for use of 5G) as key enablers for facilitating various levels of partial
614 autonomy within system of systems, which responds to remote instructions and signals, without
615 lag (latency may be fatal, for example in remote surgery or for pedestrians vs autonomous cars).
616

617   The excruciating struggle to extract information from sensor data (if there is information in
618 the data) is an indication that unleashing knowledge from information may be a mirage. The
619 anticipated evolution of data-science to knowledge-science is the central thrust of knowledge-
620 informed decision as a service (KIDS), an aspirational idea which may not be addressed by the use
621 of current tools and contemporary thinking. The broad spectrum of "data-informed" approaches
622 will vary by use cases, from simpler instances SNAPS may be a first step. For complex expectations,
623 DIDA'S may step in. Beyond DIDA'S we may be entering the domain of the unknown unknowns.
624

625   The mention of "unknown unknowns" cannot be left metaphorically hanging from the
626 rhetorical cliff of conclusion. The academic incentive to pontificate on this topic is inextricably
627 linked to the keystone factor of **knowledge discovery**. Unless we know what data to connect, the
628 extraction of information will be incorrect, insipid or incomplete, at best. Low level information is
629 akin to "the same and not the same" [92] where "whiskers" (outliers) far outweigh the central
630 substance (interquartile range) in a box plot. Using this information as knowledge bricks will be
631 tantamount to building a virtual house out of hollow cards.
632

633   Hence, knowledge discovery cannot be treated as a separate topic when discussing sensors
634 and sensor data. Without discovering the context and relevance of data to the bigger picture, the
635 outcome will remain narrow, retarded and half-baked. Sensors will be impotent without tools to
636 extract value from data fusion. This discussion is not peripheral but central, to all sensors.
637

638   The growing ubiquity of sensors, which are increasingly woven into almost every facet of our
639 daily lives, makes it imperative that sensor scientists consider the data science impact of their work.
640 Data scientists must take a closer look at sensor data and sensor engineering, to ask the correct
641 questions. Tools for knowledge discovery are not in short supply [93][94] but the rate limiting

642 factor preventing the diffusion of these tools are rooted in their complexity, lack of standards and
643 few common open platforms. If "open platforms" emerge, the race to adapt and adopt will not be
644 determined by its success due to technological strength or computational excellence. Rather the
645 economics of technology may be the single most important criteria which will influence and
646 determine feasibility [95][96] of mass adoption, the latter, in turn, will reduce cost of adoption due
647 to economies of scale. The latter is the catalyst to enable these benefits to reach the impoverished
648 communities and serve as a tool for democratization of data for the next billion users.
649
650   We are on the brink of change, albeit at a glacial pace. The advances in search and discovery of
651 data and information, are likely to improve if we can use tools based on graph theoretic
652 approaches, to establish relationships and dependencies between objects and subjects [97][98].
653 Currently, the tools from graph theory are few and far between, yet it is already being aggrandized
654 and referred to as knowledge graphs. Irrespective of the descriptive semantics and the professional
655 vernacular, knowledge graphs are potentially useful for knowledge discovery.
656
657   Scientific progress relies on the confluence of efficient assimilation of existing knowledge in
658 order to minimize re-invention. It appears that the graph theoretic approach may lead to the
659 creation of knowledge graph tools to plumb the depths of unknown unknowns, where catalysts for
660 scientific breakthroughs often reside. By using relationships between materials and their properties
661 [99] which may be organized as graphs with edges and nodes, a research team has developed a
662 methodology which may lend itself to exploring non-obvious relationships [100]. The authors are
663 incisive to point out an entity-relationship mode which remains the "bread and butter" of context-
664 awareness. The latter, in terms of knowledge representation, has evolved as RDF, the resource
665 description framework standard [101] and is a more general model of entities (nodes) and
666 relationships, albeit incomplete. Thus, the paper by Tshitoyan *et al* [99] provides tangible grounds
667 to extrapolate the methodology from material science to other domains. It strengthens the notion
668 that developing and implementing knowledge graph tools may aid in unleashing new ideas, reveal
669 unknown unknowns and enable context-aware knowledge discovery.
670
671   In another vein, a leap of progress was made when an unusual strategy was adopted to
672 organize and "number" the content (names, relationships from material science) in "preparation"
673 to use the tools of machine learning (ML) for analyzing the information and exploring
674 relationships [102][103][104][105]. Nandy *et al* **broke existing rules** and enabled analytical tools
675 to plumb the depths of the associations, generally not discovered by traditional rule-based
676 methods. It is a clear indication that the next step is to establish semantic associations [106]. We
677 propose an in-depth exploration of the tools that evolved from these materials science papers,
678 which *combined* data science with domain knowledge to explore unchartered waters [107]. The
679 principles behind these approaches may be in short supply (missing) in the domain of sensor
680 chemistry and engineering. The plethora of physico-chemical characteristics that govern the
681 attributes of sensors has little (if any) fundamental semantic association with basic science
682 (structure and properties of the molecules), which dictates functional interactions and
683 efficiency/efficacy.

This **DRAFT** copy of SNAPS is the 1st paper in a series of papers aiming to discuss *ART in SUGAR SNAPS PEAS for KIDS*

684   Interrelationships and dependencies are key to function, yet the reductionist approach must
685   view components in isolation, for the sake of operational convenience. The ability to synthesize
686   any graph theoretic approach (knowledge graphs) will depend on our ability to establish
687   connections between nodes and sub-nodes. The ability to organize the knowledge in its semantic
688   context (classification, taxonomy) may determine if we may extract *meaningful* relationships,
689   between distant, apparently unrelated nodes/sub-nodes. The ability to adopt and use the semantic
690   associations within the computational system will determine if we can use data related tools (for
691   example, machine learning, artificial neural nets) to discover relationships yet to be discovered.
692

693   Domain knowledge experts may help create nomological networks for sensor science [108]
694   perhaps resembling taxonomic trees or biochemical cycles [109]. In the next step, basic sensor
695   science must converge with the principles and practice of engineering, by learning from system
696   dynamics [110]. From the user perspective, the ability to use natural language queries [111] may be
697   beneficial. The latter unleashes the potential for neural machine reading comprehension [112].
698   However, existing tools [113][114][115][116][117][118][119][120][121] are a smörgåsbord of
699   word associations/representations [122] which may often use variations of nearest neighbor
700   principles developed circa 1000 AD [123][124]. Although pockets of evolution to elementary
701   semantic data type detection is in progress [125], almost all tools lack digital semantics [126].
702   Knowledge discovery, knowledge graphs and graph neural networks [127] will remain reasonably
703   impotent without the ability to use semantic associations, for example, semantic data catalogs
704   [128]. Thus, there appears to be a vast chasm between sensor science and engineering and data
705   science, which is in need of convergence, perhaps through a global group effort. The rejection of
706   this idea [129] may serve as a badge of honor to those who may grasp its significance.
707

708   The need for semantic tools is old news [130] but new research at the *nexus* of natural
709   language processing, linguistics, semantics, and science, may fuel trans-disciplinary convergence
710   necessary to advance knowledge discovery. Broad spectrum dissemination of this knowledge using
711   simple and tangible user interfaces may be crucial [131]. Knowledge discovery appears to be at the
712   heart of sensor research and sensor engineering, if we wish to extract value from sensors, already
713   deployed in enterprises, and aspire to deploy sensors as global public goods, for the next trillion
714   use-cases in the global supply chain [132]. For global supply chain and value network dynamics,
715   the "sense" table principle (circa 1997) using tangible user interfaces (TUI) is an useful idea, that
716   can display how data can change decisions in near real-time ([133][134][135][136][137] [138]).
717   However, TUI applications remains unexplored, in practice. Synergistic integration of tools based
718   on tangible user interfaces may be useful to simulate or display real-world potential for dynamic
719   changes and variant configurations (alternatives, options) if the operating logic of these TUI tools
720   are informed by (guided, advised, shaped) the convergence of semantically relevant curated data
721   (perishability, time sensitivity, time series), location-aware contextual information and meaningful
722   correlations which may uncover cryptic relationships (obvious, non-obvious, reinforced, learned),
723   which may influence the outcome or expectation or prediction. Information-informed tangible
724   user interface tools (iiTUIT) may be another higher order plane for combinatorial thinking and a
725   task for a new breed of creative thinkers who may invest in exploring *confluence* of ideas.

726 **8. Concluding Remarks**

727   Ubiquity of sensor data makes it crucial that sensor data offers semantically sensible
728   value, preferably at every instance, starting with less complex point solutions, referred to as
729   SNAPS, in this discussion. SNAPS and other layers in the layer cake (0-6) illustrated as steps
730   are not discrete stages but represents tool sets and or decision sub-platforms which can
731   combine, re-combine and co-exist in various forms. For example, artificial reasoning tools
732   (ART) is a sub-platform where the output may use various types of logic or rules which may
733   guide/shape rational/irrational decisions. ART is not only its own layer but may be viewed as a
734   mobile layer (repertoire of tools) which can work with various other layers to modulate the
735   output (decision). For example, SUGAR (simple user guided activity report) only delivers raw
736   data based on user command. If the output from SUGAR feeds ART, the tools may offer a
737   preliminary processing of data generated by SUGAR and offer an output (decision support)
738   which is a level above SUGAR (yet, not SNAPS). The "empty boxes" (indicated on the right)
739   in the illustration represents these potential combinations which may be case specific and
740   offers room for innovation as the evolution of decision science continues. ART may be one
741   occupant of these "boxes" if necessary, in a certain scenario (use case). ART may work with
742   any layer and in any stage (in any "box"). In this bio-inspired "McClintock"-esque paradigm
743   [139], ART (or something similar or equivalent) may be viewed as a cross-pollinator of
744   SUGAR, SNAPS, PEAS (hence "jumping genes" may be a suitable analogy [140] as well as the
745   related "mobile transposon" concept [141]). The threads of these ideas advocate the potential
746   of cross-platform mobile elements (tools) which may be designed to act as dynamic cross-
747   fertilizing "influenza" {(the term "influenza" is Italian for "influence" and used [142] since
748   circa 1357 (circa 1504 in English) to describe the illness, due to its perceived association with
749   cold weather (influenze di freddo)} agents capable of improving and/or influencing, as well as
750   linking, various domains of decision support systems.
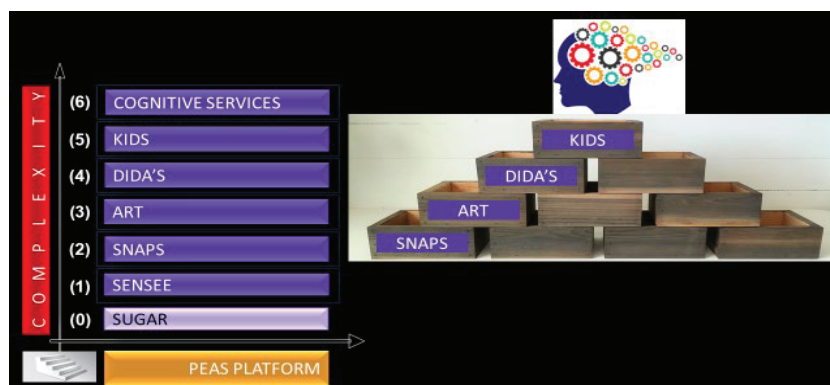
751

752   **Figure 9**. The challenge for a new breed of thinkers is to pave the path with value
753   extracted from data and information from SUGAR (simple user guided activity report) to
754   SERVICES (layer 6), and the layers and sub-layers, which may be necessary or
755   mandatory but still remains to be discovered and/or identified. In the distant future,
756   perhaps, someday, the system may legitimately claim to integrate the use of cognitive
757   tools even beyond knowledge (KIDS, knowledge informed decision as a service) to
758   introduce and promote bold new frontiers in the realm of decision science.

This **DRAFT** copy of SNAPS is the 1st paper in a series of papers aiming to discuss *ART in SUGAR SNAPS PEAS for KIDS*

759 **Author Contributions:**

760 Conceptualization: ESM, SPAD

761 Data curation: YR

762 Formal analysis: YR

763 Funding acquisition: ESM

764 Formatting: ESM, SB

765 Investigation: VM, NC

766 Methodology: ESM

767 Project administration: ESM, DV

768 Resources: ESM, DV

769 Software: VM, NC, YR, SB

770 Supervision: ESM

771 Visualization: ESM, SPAD

772 Writing – original draft: ESM, SPAD

773 Writing – review and editing: ALL

774

778

779 **Conflicts of Interest:** "The authors declare no conflict of interest."

780

781 **References**

782 1. Wang, J., 2001. Glucose biosensors: 40 years of advances and challenges. Electroanalysis: An
783   International Journal Devoted to Fundamental and Practical Aspects of
784   Electroanalysis, 13(12), pp.983-988.

785 2. Wang, B. and Anzai, J.I., 2015. Recent progress in lectin-based biosensors. *Materials*, *8*(12),
786   pp.8590-8607..

787 3. Song, Y., Luo, Y., Zhu, C., Li, H., Du, D. and Lin, Y., 2016. Recent advances in electrochemical
788   biosensors based on graphene two-dimensional nanomaterials. *Biosensors and*
789   *Bioelectronics*, *76*, pp.195-212.

790 4. Yang, L. and Bashir, R., 2008. Electrical/electrochemical impedance for rapid detection of
791   foodborne pathogenic bacteria. *Biotechnology advances*, *26*(2), pp.135-150.

792 5. Brolo, A.G., 2012. Plasmonics for future biosensors. *Nature Photonics*, *6*(11), p.709.

793 6. Scheller, F.W., Yarman, A., Bachmann, T., Hirsch, T., Kubick, S., Renneberg, R., Schumacher,
794   S., Wollenberger, U., Teller, C. and Bier, F.F., 2013. Future of biosensors: a personal view.
795   In *Biosensors Based on Aptamers and Enzymes* (pp. 1-28). Springer, Berlin, Heidelberg.

796 7. Li, R., Feng, Y., Pan, G. and Liu, L., 2019. Advances in molecularly imprinting technology for
797   bioanalytical applications. *Sensors*, *19*(1), p.177.

798 8. Murphy, L., 2006. Biosensors and bioelectrochemistry. *Current opinion in chemical*
799   *biology*, *10*(2), pp.177-184.

800 9. Bogue, R., 2009. Nanosensors: a review of recent research. *Sensor Review*, *29*(4), pp.310-315.

801 10. Li, C., Bai, H. and Shi, G., 2009. Conducting polymer nanomaterials: electrosynthesis and
802   applications. *Chemical Society Reviews*, *38*(8), pp.2397-2409.

803 11. Bonanni, A., Loo, A.H. and Pumera, M., 2012. Graphene for impedimetric biosensing. *TrAC*
804   *Trends in Analytical Chemistry*, *37*, pp.12-21.

805 12. Zhu, C., Yang, G., Li, H., Du, D. and Lin, Y., 2014. Electrochemical sensors and biosensors
806   based on nanomaterials and nanostructures. *Analytical chemistry*, *87*(1), pp.230-249.

807 13. Ekinci, K.L., 2005. Electromechanical transducers at the nanoscale: actuation and sensing of
808   motion in nanoelectromechanical systems (NEMS). *small*, *1*(8-9), pp.786-797.

809 14. Collins, B.E. and Anslyn, E.V., 2007. Pattern-Based Peptide Recognition. *Chemistry–A*
810   *European Journal*, *13*(17), pp.4700-4708.

811 15. Turner, A.P., 2013. Biosensors: sense and sensibility. *Chemical Society Reviews*, *42*(8),
812   pp.3184-3196.

813 16. Thakare, V. and Khire, G., 2014. Role of emerging technology for building smart hospital
814   information system. *Procedia Economics and Finance*, *11*, pp.583-588.

815 17. Quesada-González, D. and Merkoçi, A., 2017. Mobile phone-based biosensing: An emerging
816   "diagnostic and communication" technology. *Biosensors and Bioelectronics*, *92*, pp.549-562.

817 18. Wei, Q., Luo, W., Chiang, S., Kappel, T., Mejia, C., Tseng, D., Chan, R.Y.L., Yan, E., Qi, H.,
818   Shabbir, F. and Ozkan, H., 2014. Imaging and sizing of single DNA molecules on a mobile
819   phone. *ACS nano*, *8*(12), pp.12725-12733.

820 19. Guner, H., Ozgur, E., Kokturk, G., Celik, M., Esen, E., Topal, A.E., Ayas, S., Uludag, Y.,
821   Elbuken, C. and Dana, A., 2017. A smartphone based surface plasmon resonance imaging
822   (SPRi) platform for on-site biodetection. *Sensors and Actuators B: Chemical*, *239*, pp.571-577.

823 20. Rowe, A.A., Bonham, A.J., White, R.J., Zimmer, M.P., Yadgar, R.J., Hobza, T.M., Honea, J.W.,
824   Ben-Yaacov, I. and Plaxco, K.W., 2011. CheapStat: an open-source,"Do-It-Yourself"
825   potentiostat for analytical and educational applications. *PloS one*, *6*(9), p.e23783.

826 21. Jenkins, D.M., Lee, B.E., Jun, S., Reyes-De-Corcuera, J. and McLamore, E.S., 2019. ABE-Stat, a
827   Fully Open-Source and Versatile Wireless Potentiostat Project Including Electrochemical
828   Impedance Spectroscopy. *Journal of The Electrochemical Society*, *166*(9), pp.B3056-B3065..

829 22. Moraru, A., Pesko, M., Porcius, M., Fortuna, C. and Mladenic, D., 2010. Using machine
830   learning on sensor data. *Journal of computing and information technology*, *18*(4), pp.341-347.

831 23. Ghahramani, Z., 2015. Probabilistic machine learning and artificial
832   intelligence. *Nature*, *521*(7553), p.452.

833 24. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D.C. and Kording, K.P., 2016. Voodoo machine
834   learning for clinical predictions. *Biorxiv*, p.059774.

835   25.   Rong, Y., Padron, A.V., Hagerty, K.J., Nelson, N., Chi, S., Keyhani, N.O., Katz, J., Datta,
836   S.P.A., Gomes, C. and McLamore, E.S., 2018. Post hoc support vector machine learning for
837   impedimetric biosensors based on weak protein–ligand interactions. *Analyst*, *143*(9),
838   pp.2066-2075.
839   26.   Koczula, K.M. and Gallotta, A., 2016. Lateral flow assays. *Essays in biochemistry*, *60*(1),
840   pp.111-120.
841   27.   Olsman, N. and Goentoro, L., 2016. Allosteric proteins as logarithmic sensors. *Proceedings of*
842   *the National Academy of Sciences*, *113*(30), pp.E4423-E4430.
843   28.   Nussinov, R., Tsai, C.J. and Ma, B., 2013. The underappreciated role of allostery in the cellular
844   network. *Annual review of biophysics*, *42*, pp.169-189.
845   29.   Taguchi, M., Ptitsyn, A., McLamore, E.S. and Claussen, J.C., 2014. Nanomaterial-mediated
846   biosensors for monitoring glucose. *Journal of diabetes science and technology*, *8*(2), pp.403-
847   411.
848   30.   Mosbah, A., Campanacci, V., Lartigue, A., Tegoni, M., Cambillau, C. and Darbon, H., 2003.
849   Solution structure of a chemosensory protein from the moth Mamestra brassicae. *Biochemical*
850   *Journal*, *369*(1), pp.39-44.
851   31.   Vámos, T., 1992. Judea pearl: Probabilistic reasoning in intelligent systems. *Decision Support*
852   *Systems*, *8*(1), pp.73-75.
853   32.   Marr, D. and Poggio, T., 1979. A computational theory of human stereo vision. *Proceedings of*
854   *the Royal Society of London. Series B. Biological Sciences*, *204*(1156), pp.301-328.
855   33.   Bhusal, N., Shrestha, S., Pote, N. and Alocilja, E., 2019. Nanoparticle-Based Biosensing of
856   Tuberculosis, an Affordable and Practical Alternative to Current Methods. *Biosensors*, *9*(1),
857   p.1.
858   34.   Gordillo-Marroquín, C., Gómez-Velasco, A., Sánchez-Pérez, H., Pryg, K., Shinners, J.,
859   Murray, N., Muñoz-Jiménez, S., Bencomo-Alerm, A., Gómez-Bustamante, A., Jonapá-Gómez,
860   L. and Enríquez-Ríos, N., 2018. Magnetic Nanoparticle-Based Biosensing Assay
861   Quantitatively Enhances Acid-Fast Bacilli Count in Paucibacillary Pulmonary
862   Tuberculosis. *Biosensors*, *8*(4), p.128.
863   35.   Wei, Q., Qi, H., Luo, W., Tseng, D., Ki, S.J., Wan, Z., Göröcs, Z., Bentolila, L.A., Wu, T.T.,
864   Sun, R. and Ozcan, A., 2013. Fluorescent imaging of single nanoparticles and viruses on a
865   smart phone. *ACS nano*, *7*(10), pp.9147-9155.
866   36.   Zheng, L., Cai, G., Wang, S., Liao, M., Li, Y. and Lin, J., 2019. A microfluidic colorimetric
867   biosensor for rapid detection of Escherichia coli O157: H7 using gold nanoparticle
868   aggregation and smart phone imaging. *Biosensors and Bioelectronics*, *124-125*, pp.143-149.
869   37.   Yetisen, A.K., Martinez-Hurtado, J.L., Garcia-Melendrez, A., da Cruz Vasconcellos, F. and
870   Lowe, C.R., 2014. A smartphone algorithm with inter-phone repeatability for the analysis of
871   colorimetric tests. *Sensors and Actuators B: Chemical*, *196*, pp.156-160.
872   38.   Lopez-Ruiz, N., Curto, V.F., Erenas, M.M., Benito-Lopez, F., Diamond, D., Palma, A.J. and
873   Capitan-Vallvey, L.F., 2014. Smartphone-based simultaneous pH and nitrite colorimetric
874   determination for paper microfluidic devices. *Analytical chemistry*, *86*(19), pp.9554-9562.
875   39.   Thompson, R.E., Larson, D.R. and Webb, W.W., 2002. Precise nanometer localization
876   analysis for individual fluorescent probes. *Biophysical journal*, *82*(5), pp.2775-2783.

877 40. Gunda, N.S.K., Gautam, S.H. and Mitra, S.K., 2019. Artificial Intelligence Based Mobile
878   Application for Water Quality Monitoring. *Journal of The Electrochemical Society*, *166*(9),
879   pp.B3031-B3035.
880 41. Vanegas, D., Patiño, L., Mendez, C., Oliveira, D., Torres, A., Gomes, C. and McLamore, E.,
881   2018. Laser Scribed Graphene Biosensor for Detection of Biogenic Amines in Food Samples
882   Using Locally Sourced Materials. *Biosensors*, *8*(2), p.42.
883 42. Yoo, S.M. and Lee, S.Y., 2016. Optical biosensors for the detection of pathogenic
884   microorganisms. *Trends in biotechnology*, *34*(1), pp.7-25.
885 43. Liu, L., Zhang, D., Zhang, Q., Chen, X., Xu, G., Lu, Y. and Liu, Q., 2017. Smartphone-based
886   sensing system using ZnO and graphene modified electrodes for VOCs detection. *Biosensors*
887   *and Bioelectronics*, *93*, pp.94-101.
888 44. Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T. and Campbell, A.T., 2010. A
889   survey of mobile phone sensing. *IEEE Communications magazine*, *48*(9), pp.140-150.
890 45. McGrath, A.P., Hilmer, K.M., Collyer, C.A., Shepard, E.M., Elmore, B.O., Brown, D.E.,
891   Dooley, D.M. and Guss, J.M., 2009. Structure and inhibition of human diamine
892   oxidase. *Biochemistry*, *48*(41), pp.9810-9822.
893 46. Babiceanu, R.F. and Seker, R., 2016. Big Data and virtualization for manufacturing cyber-
894   physical systems: A survey of the current status and future outlook. *Computers in Industry*, *81*,
895   pp.128-137.
896 47. C. Wu, R. Buyya, and K. Ramamohanarao, 2016 . "Big Data Analytics = Machine Learning +
897   Cloud Computing," in Buyya, R., Calheiros, R.N. and Dastjerdi, A.V. eds. *Big Data: Principles*
898   *and Paradigms*. Elsevier, pp. 1-27.
899 48. Ravi, K., Khandelwal, Y., Krishna, B.S. and Ravi, V., 2018. Analytics in/for cloud-an
900   interdependence: A review. *Journal of Network and Computer Applications*, *102*, pp.17-37.
901 49. J Heer, J. and Perer, A., 2014. Orion: A system for modeling, transformation and visualization
902   of multidimensional heterogeneous networks. *Information Visualization*, *13*(2), pp.111-133.
903 50. Kim, J., Levy, E., Ferbrache, A., Stepanowsky, P., Farcas, C., Wang, S., Brunner, S., Bath, T.,
904   Wu, Y. and Ohno-Machado, L., 2014. MAGI: a Node. js web service for fast microRNA-Seq
905   analysis in a GPU infrastructure. *Bioinformatics*, *30*(19), pp.2826-2827.
906 51. Nolte, H., MacVicar, T.D., Tellkamp, F. and Krüger, M., 2018. Instant Clue: a software suite
907   for interactive data visualization and analysis. *Scientific reports*, *8*(1), p.12648.
908 52. Ko, G., Kim, P.G., Yoon, J., Han, G., Park, S.J., Song, W. and Lee, B., 2018. Closha:
909   bioinformatics workflow system for the analysis of massive sequencing data. *BMC*
910   *bioinformatics*, *19*(1), p.43.
911 53. Shang, Z., Zgraggen, E., Buratti, B., Kossmann, F., Eichmann, P., Chung, Y., Binnig, C., Upfal,
912   E. and Kraska, T., 2019, June. Democratizing data science through interactive curation of ml
913   pipelines. In *Proceedings of the 2019 International Conference on Management of Data* (pp.
914   1171-1188). ACM.
915 54. Binnig, C., Buratti, B., Chung, Y., Cousins, C., Kraska, T., Shang, Z., Upfal, E., Zeleznik, R. and
916   Zgraggen, E., 2018, June. Towards interactive curation & automatic tuning of ml pipelines.
917   In *Proceedings of the Second Workshop on Data Management for End-To-End Machine*
918   *Learning* (p. 1). ACM.

919 55. Chung, Y., Servan-Schreiber, S., Zgraggen, E. and Kraska, T., 2018. Towards Quantifying
920   Uncertainty in Data Analysis & Exploration. *IEEE Data Eng. Bull.*, *41*(3), pp.15-27.
921 56. Tou, J.T. and Gonzalez, R.C., 1972. Automatic recognition of handwritten characters via
922   feature extraction and multi-level decision. *International Journal of Computer & Information*
923   *Sciences*, *1*(1), pp.43-65.
924 57. Zeng, W., Meng, X., Yang, C. and Huang, L., 2006. Feature extraction for online handwritten
925   characters using Delaunay triangulation. *Computers & Graphics*, *30*(5), pp.779-786.
926 58. COMSOL, "Introduction to COMSOL Multiphysics 5.3," *Manual*, 2014.
927 59. P. Roger W. Pryor, *Multiphysics Modeling Using COMSOL®: A First Principles Approach*. 2009.
928 60. Hamada, M. and Sato, S., 2010, June. Lego NXT as a learning tool. In *Proceedings of the*
929   *fifteenth annual conference on Innovation and technology in computer science education* (pp.
930   321-321). ACM.
931 61. Datta, S.P.A., 2017. Emergence of Digital Twins-Is this the march of reason?. *Journal of*
932   *Innovation Management*, *5*(3), pp.14-33.
933 62. Meystel, A., 1989. Intelligent control: A sketch of the theory. *Journal of Intelligent & Robotic*
934   *Systems*, *2*(2), pp.97-107.
935 63. Stephanopoulos, G. and Han, C., 1996. Intelligent systems in process engineering: A
936   review. *Computers & Chemical Engineering*, *20*(6-7), pp.743-791.
937 64. Meystel, A., 1993, August. Architectures for intelligent control systems: The science of
938   autonomous intelligence. In *Proceedings of 8th IEEE International Symposium on Intelligent*
939   *Control* (pp. 42-48). IEEE.
940 65. Antsaklis, P.J., Passino, K.M. and Wang, S.J., 1991. An introduction to autonomous control
941   systems. *IEEE Control Systems Magazine*, *11*(4), pp.5-13.
942 66. Albaladejo, C., Soto, F., Torres, R., Sánchez, P. and López, J.A., 2012. A low-cost sensor buoy
943   system for monitoring shallow marine environments. *Sensors(Switzerland)*, *12*(7), pp.9613-
944   9634.
945 67. Pardee, A.B., Jacob, F. and Monod, J., 1959. The genetic control and cytoplasmic expression of
946   "inducibility" in the synthesis of β-galactosidase by E. coli. *Journal of Molecular Biology*, *1*(2),
947   pp.165-178.
948 68. Quinlan, J.R., 1986. Induction of decision trees. *Machine learning*, *1*(1), pp.81-106.
949 69. "Introduction to Decision Trees," 2014.
950 70. Forstmann, B.U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D.Y., Ridderinkhof, K.R.
951   and Wagenmakers, E.J., 2008. Striatum and pre-SMA facilitate decision-making under time
952   pressure. *Proceedings of the National Academy of Sciences*, *105*(45), pp.17538-17542.
953 71. Gold, C., Damböck, D., Lorenz, L. and Bengler, K., 2013, September. "Take over!" How long
954   does it take to get the driver back into the loop?. In *Proceedings of the Human Factors and*
955   *Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1938-1942). Sage CA: Los Angeles,
956   CA: SAGE Publications..
957 72. Cabri, G., Leonardi, L. and Zambonelli, F., 2000. MARS: A programmable coordination
958   architecture for mobile agents. *IEEE Internet Computing*, *4*(4), pp.26-35.
959 73. E. Jean, 2011. "Sensor network interoperability and reconfiguration through mobile agents,".

960 74. Fielding, R.T. and Taylor, R.N., 2000. *Architectural styles and the design of network-based*
961 *software architectures*. Doctoral dissertation: University of California, Irvine.
962 75. Willinger, W., Paxson, V., Riedi, R.H. and Taqqu, M.S., 2003. Long-range dependence and
963 data network traffic. *Theory and applications of long-range dependence*, pp.373-407.
964 76. Vélez-Torres, I., Vanegas, D.C., McLamore, E.S. and Hurtado, D., 2018. Mercury pollution
965 and artisanal gold mining in Alto Cauca, Colombia: woman's perception of health and
966 environmental impacts. *The Journal of Environment & Development*, 27(4), pp.415-444.
967 77. Abdelbasir, S.M., El-Sheikh, S.M., Morgan, V.L., Schmidt, H., Casso-Hartmann, L.M.,
968 Vanegas, D.C., Velez-Torres, I. and McLamore, E.S., 2018. Graphene-anchored cuprous oxide
969 nanoparticles from waste electric cables for electrochemical sensing. *ACS Sustainable*
970 *Chemistry & Engineering*, 6(9), pp.12176-12186.
971 78. Nakazawa, K., Nagafuchi, O., Kawakami, T., Inoue, T., Yokota, K., Serikawa, Y., Cyio, B. and
972 Elvince, R., 2016. Human health risk assessment of mercury vapor around artisanal small-
973 scale gold mining area, Palu city, Central Sulawesi, Indonesia. *Ecotoxicology and*
974 *environmental safety*, 124, pp.155-162.
975 79. Saleem, M., Iqbal, J. and Shah, M.H., 2014. Dissolved concentrations, sources, and risk
976 evaluation of selected metals in surface water from mangla lake, Pakistan. *The Scientific World*
977 *Journal*, 2014.
978 80. Xiao, J., Wang, L., Deng, L. and Jin, Z., 2019. Characteristics, sources, water quality and health
979 risk assessment of trace elements in river water and well water in the Chinese Loess
980 Plateau. *Science of the Total Environment*, 650, pp.2004-2012.
981 81. Takabe, Y., Tsuno, H., Nishimura, F., Tanii, N., Maruno, H., Tsurukawa, M., Suzuki, M. and
982 Matsumura, C., 2012. Bioaccumulation and primary risk assessment of persistent organic
983 pollutants with various bivalves. *Water Science and Technology*, 66(12), pp.2620-2629.
984 82. U. S. E. P. a Oppt, "Quantitative Risk Assessment Calculations," *EPA Sustain. Futur. Framew.*
985 *Man. 2012 EPA-748-B12-001*, 2012.
986 83. Smith, A.H., Lingas, E.O. and Rahman, M., 2000. Contamination of drinking-water by arsenic
987 in Bangladesh: a public health emergency. *Bulletin of the World Health Organization*, 78,
988 pp.1093-1103.
989 84. Van Geen, A., Trevisani, M., Immel, J., Jakariya, M., Osman, N., Cheng, Z., Gelman, A. and
990 Ahmed, K.M., 2006. Targeting low-arsenic groundwater with mobile-phone technology in
991 Araihazar, Bangladesh. *Journal of health, population, and nutrition*, 24(3), p.282.
992 85. Wan, X., Volpetti, F., Petrova, E., French, C., Maerkl, S.J. and Wang, B., 2019. Cascaded
993 amplifying circuits enable ultrasensitive cellular sensors for toxic metals. *Nature chemical*
994 *biology*, 15(5), p.540.
995 86. Haque, F., Ball, R.L., Khatun, S., Ahmed, M., Kache, S., Chisti, M.J., Sarker, S.A., Maples, S.D.,
996 Pieri, D., Korrapati, T.V. and Sarnquist, C., 2017. Evaluation of a smartphone decision-
997 support tool for diarrheal disease management in a resource-limited setting. *PLoS neglected*
998 *tropical diseases*, 11(1), p.e0005290.
999 87. Hills, K.D., Oliveira, D.A., Cavallaro, N.D., Gomes, C.L. and McLamore, E.S., 2018. Actuation
1000 of chitosan-aptamer nanobrush borders for pathogen sensing. *Analyst*, 143(7), pp.1650-1661.

88. Shirude, S.B. and Kolhe, S.R., 2018. Agent-Based Architecture for Developing Recommender System in Libraries. In *Knowledge Computing and its Applications* (pp. 157-181). Springer, Singapore.

89. Ellul, J. (1964). The technological society. New York: Knopf.

90. Subcommittee: F29.21 (ASTM), "ASTM F2761 - 09(2013) Medical Devices and Medical Systems - Essential safety requirements for equipment comprising the patient-centric integrated clinical environment (ICE)," 2013.

91. Makary, M.A. and Daniel, M., 2016. Medical error—the third leading cause of death in the US. *Bmj*, *353*, p.i2139.

92. R. Hoffmann and F. J. Dyson, "The Same and Not the Same" *Phys. Today*, 2008.

93. W. E. McCarthy, "Knowledge Representation: Logical, Philosophical, and Computational Foundations," *Account. Rev.*, 2002.

94. Bergman, M.K., 2018. *A Knowledge Representation Practionary: Guidelines Based on Charles Sanders Peirce*. Springer.

95. P. A. David, and G. Wright (2012) General Purpose Technologies and Surges in Productivity: Historical Reflections on the Future of the ICT Revolution, in *The Economic Future in Historical Perspective*, 2012.

96. *The Economic Future in Historical Perspective*. 2012.

97. Bullmore, E. and Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, *10*(3), p.186.

98. Fornito, A., Zalesky, A. and Breakspear, M., 2013. Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage*, *80*, pp.426-444.

99. Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K.A., Ceder, G. and Jain, A., 2019. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*, *571*(7763), p.95.

100. J. Jonas, 2005. Non-Obvious Relationship Awareness (NORA), in *IBM Entity Analytic Solutions*, pp. 1–78.

101. G. Bartolomeo and T. Kovacikova. (2013). "Resource Description Framework", in G. Bartolomeo and T. Kovacikova. *Identification and Management of Distributed Data*, Boca Raton: CRC Press, pp. 205-211.

102. Janet, J.P., Chan, L. and Kulik, H.J., 2018. Accelerating chemical discovery with machine learning: simulated evolution of spin crossover complexes with an artificial neural network. *The journal of physical chemistry letters*, *9*(5), pp.1064-1071.

103. Janet, J.P. and Kulik, H.J., 2017. Predicting electronic structure properties of transition metal complexes with neural networks. *Chemical science*, *8*(7), pp.5137-5152.

104. Janet, J.P. and Kulik, H.J., 2017. Resolving transition metal chemical space: Feature selection for machine learning and structure–property relationships. *The Journal of Physical Chemistry A*, *121*(46), pp.8939-8954.

105. Nandy, A., Duan, C., Janet, J.P., Gugler, S. and Kulik, H.J., 2018. Strategies and software for machine learning accelerated discovery in transition metal chemistry. *Industrial & Engineering Chemistry Research*, *57*(42), pp.13973-13986.

1042   106. Barati, M., Bai, Q. and Liu, Q., 2017. Mining semantic association rules from RDF
1043   data. *Knowledge-Based Systems*, *133*, pp.183-196.
1044   107. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H. and Kang, J., 2019. Biobert: pre-trained
1045   biomedical language representation model for biomedical text mining. *arXiv preprint*
1046   *arXiv:1901.08746*.
1047   108. Cronbach, L.J. and Meehl, P.E., 1955. Construct validity in psychological tests. *Psychological*
1048   *bulletin*, *52*(4), p.281.
1049   109. Krebs, H.A. and Johnson, W.A., 1937. The role of citric acid in intermediate metabolism in
1050   animal tissues. *Enzymologia*, *4*, pp.148-156.
1051   110. Forrester, J.W., 1995. The beginning of system dynamics. *McKinsey Quarterly*, pp.4-17.
1052   111. Guo, J., Zhan, Z., Gao, Y., Xiao, Y., Lou, J.G., Liu, T. and Zhang, D., 2019. Towards Complex
1053   Text-to-SQL in Cross-Domain Database with Intermediate Representation. *arXiv preprint*
1054   *arXiv:1905.08205*.
1055   112. Liu, S., Zhang, X., Zhang, S., Wang, H. and Zhang, W., 2019. Neural machine reading
1056   comprehension: Methods and trends. *Applied Sciences*, *9*(18), p.3698.
1057   113. Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep
1058   bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
1059   114. Shen, Y., Huang, P.S., Gao, J. and Chen, W., 2017, August. Reasonet: Learning to stop reading
1060   in machine comprehension. In *Proceedings of the 23rd ACM SIGKDD International*
1061   *Conference on Knowledge Discovery and Data Mining* (pp. 1047-1055). ACM.
1062   115. Wang, W., Yang, N., Wei, F., Chang, B. and Zhou, M., 2017, July. Gated self-matching
1063   networks for reading comprehension and question answering. In *Proceedings of the 55th*
1064   *Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp.
1065   189-198).
1066   116. Tan, C., Wei, F., Yang, N., Du, B., Lv, W. and Zhou, M., 2017. S-net: From answer extraction
1067   to answer generation for machine reading comprehension. *arXiv preprint arXiv:1706.04815*.
1068   117. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. and Le, Q.V., 2019. XLNet:
1069   Generalized Autoregressive Pretraining for Language Understanding. *arXiv preprint*
1070   *arXiv:1906.08237*.
1071   118. Wang, Y., Wang, S., Tang, J., O'Hare, N., Chang, Y. and Li, B., 2016. Hierarchical attention
1072   network for action recognition in videos. *arXiv preprint arXiv:1607.06416*.
1073   119. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. and Hovy, E., 2016, June. Hierarchical
1074   attention networks for document classification. In *Proceedings of the 2016 conference of the*
1075   *North American chapter of the association for computational linguistics: human language*
1076   *technologies* (pp. 1480-1489).
1077   120. Jain, S. and Wallace, B.C., 2019. Attention is not explanation. *arXiv preprint*
1078   *arXiv:1902.10186*.
1079   121. Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q. and Shen, D., 2017, September.
1080   Medical image synthesis with context-aware generative adversarial networks. In *International*
1081   *Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 417-425).
1082   Springer, Cham.

1083   122. Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word
1084   representations in vector space. *arXiv preprint arXiv:1301.3781*.
1085   123. Chen, G.H. and Shah, D., 2018. Explaining the success of nearest neighbor methods in
1086   prediction. *Foundations and Trends® in Machine Learning*, *10*(5-6), pp.337-588.
1087   124. Pelillo, M., 2014. Alhazen and the nearest neighbor rule. *Pattern Recognition Letters*, *38*,
1088   pp.34-37.
1089   125. Hulsebos, M., Hu, K., Bakker, M., Zgraggen, E., Satyanarayan, A., Kraska, T., Demiralp, Ç.
1090   and Hidalgo, C., 2019. Sherlock: A Deep Learning Approach to Semantic Data Type Detection
1091   In *25TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA*
1092   *MINING*, Anachorage, Alaska, 4-8 August.
1093   *126.* Datta, S.P.A., 2007. Unified theory of relativistic identification of information in a systems
1094   age: Proposed convergence of unique identification with syntax and semantics through
1095   Internet protocol version 6.
1096   127. Xu, K., Li, J., Zhang, M., Du, S.S., Kawarabayashi, K., & Jegelka, S., 2019. What Can Neural
1097   Networks Reason About? *ArXiv, abs/1905.13211*.
1098   128. www.schema.org
1099   129. Borrell, B., 2010. Foundations-Nature Rejects Krebs's Paper, 1937. *Scientist*, *24*(3), p.88.
1100   130. Datta, S., 2008. Potential for improving decision support catalysed by semantic
1101   interoperability between systems.
1102   131. Patten, J., Ishii, H., Hines, J. and Pangaro, G., 2001, March. Sensetable: a wireless object
1103   tracking platform for tangible user interfaces. In *Proceedings of the SIGCHI conference on*
1104   *Human factors in computing systems.* ACM, pp. 253-260.
1105   132. Datta, S., Betts, B., Dinning, M., Erhun, F., Gibbs, T., Keskinocak, P., Li, H., Li, M. and
1106   Samuels, M., 2004. Adaptive value networks. In *Evolution of supply chain management* (pp. 3-
1107   67). Springer, Boston, MA.
1108   133. Ullmer, B. and Ishii, H., 1997. The metaDESK: models and prototypes for tangible user
1109   interfaces. In *Proceedings of Symposium on User Interface Software and Technology (UIST???*
1110   *97), ACM*.
1111   134. Ishii, H., 2004, lecture notes, Tangible Media Group MIT Media Laboratory, Tokyo, delivered
1112   February 2004.
1113   135. Maquil, V., Leopold, U., De Sousa, L.M., Schwartz, L. and Tobias, E., 2018. Towards a
1114   framework for geospatial tangible user interfaces in collaborative urban planning. *Journal of*
1115   *Geographical Systems*, *20*(2), pp.185-206.
1116   136. Ullmer, B. and Ishii, H., 2000. Emerging frameworks for tangible user interfaces. *IBM systems*
1117   *journal*, *39*(3.4), pp.915-931.
1118   137. Ullmer, B.A., 1997. *Models and mechanisms for tangible user interfaces* (Master Thesis,
1119   Massachusetts Institute of Technology).
1120   138. https://tangible.media.mit.edu/papers/
1121   139. Pray, L. & Zhaurova, K. 2008. Barbara McClintock and the discovery of jumping genes
1122   (transposons). *Nature Education* 1(1):169.

1123 140. Creighton, H.B. and McClintock, B., 1931. A correlation of cytological and genetical crossing-
1124   over in Zea mays. *Proceedings of the National Academy of Sciences of the United States of*
1125   *America*, *17*(8), p.492.
1126 141. Ravindran, S., 2012. Barbara McClintock and the discovery of jumping genes. *Proceedings of*
1127   *the National Academy of Sciences*, *109*(50), pp.20198-20199.
1128 142. Prince, T. 2019, 'The Definitive History of the Flu', *Champagne remedies, sneezing ferrets, and*
1129   *thousands of years of havoc*. https://elemental.medium.com/the-definitive-history-of-the-flu-
1130   b975432f9fc5

1131

1132

This **DRAFT** copy of SNAPS is the 1st paper in a series of papers aiming to discuss *ART in SUGAR SNAPS PEAS for KIDS*

Beginnings | 1st Neural Winter | 2nd Neural Winter | GPU Era

| Thresholded Logic Unit | Perceptron | Adaline | XOR Problem | Multilayer Backprop | CNNs | LSTMs | SVMs | Deep Nets | Alex Net |

| 1943 | 1957 | 1960 | 1969 | 1982 | 1986 | 1989 | 1997 | 1995 | 2006 | 2012 |

1940 1950 1960 1970 1980 1990 2000 2010

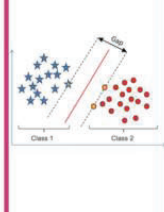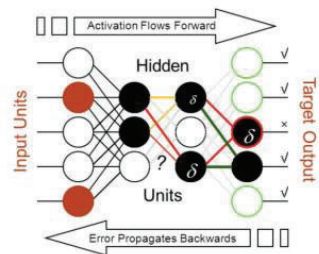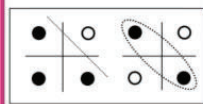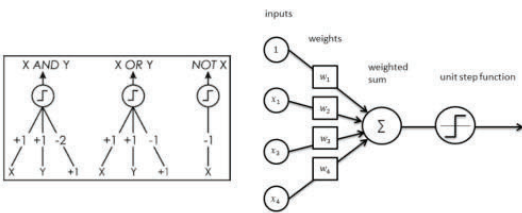S. McCulloch - W. Pitts | R. Rosenblatt | B. Widrow - M. Hoff | M. Minsky - S. Papert | P. Werbos | D. Rumelhart - G. Hinton - R. Williams | Y. Lecun - J. Schmidhuber | C. Cortes - V. Vapnik | R. Salakhutdinov - J. Hinton - A. Krizhevsky - I. Sutskever
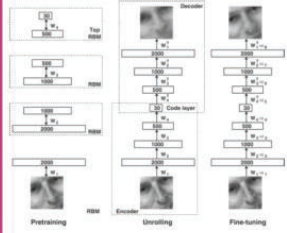
review

Sensor-as-a-Service: Convergence of Sensor Analytic Point Solutions (SNAPS) and Pay-A-Penny-Per-Use (PAPPU) Paradigm as a Catalyst for Democratization of Healthcare in Underserved Communities

Victoria Morgan[1], Lisseth Casso-Hartman[2,3], David Bahamon-Pinzon[4], Kelli McCourt[4], Robert G. Hjort[5], Sahar Bahramzadeh[6], Irene Velez-Torres[2,3], Eric McLamore[1], Carmen Gomes[5], Evangelyn C. Alocilja[7,8], Nirajan Bhusal[7,9,10], Sunaina Shrestha[10], Nisha Pote[10], Ruben Kenny Briceno[11,12,13,7], Shoumen Palit Austin Datta[1,14,15,16], Diana C. Vanegas[4,3*]

[1] Agricultural and Biological Engineering, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, 32611, USA

[2] Natural Resources and Environmental Engineering, Universidad del Valle, Cali, Colombia

[3] Interdisciplinary Group for Biotechnological Innovation and Ecosocial Change -BioNovo, Universidad del Valle, Cali, Colombia

[4] Biosystems Engineering, Department of Environmental Engineering and Earth Sciences, Clemson University, Clemson, SC 29631, USA

[5] Mechanical Engineering, Iowa State University, Ames, IA, 50011, USA

[6] School of Computer Engineering, Azad University, Science and Research Branch, Saveh, Iran

[7] Global Alliance for Rapid Diagnostics, Michigan State University, East Lansing, MI 48824, USA

[8] Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA

[9] School of Medical Sciences, Kathmandu University, Kathmandu, Nepal

[10] Dhulikhel Hospital, Kathmandu University, Kavrepalanchok, Nepal

[11] Instituto de Investigacion en Ciencia y Tecnologia, Universidad Cesar Vallejo, Trujillo 13100, Peru

[12] Hospital Victor Lazarte Echegaray, Trujillo 13100, Peru

[13] Institute for Global Health, Michigan State University, East Lansing, MI 48824, USA

[14] MIT Auto-ID Labs, Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139. USA

[15] MDPnP Interoperability and Cybersecurity Labs, Biomedical Engineering Program, Department of Anesthesiology, Massachusetts General Hospital, Harvard Medical School, 65 Landsdowne Street, Cambridge, MA 02139, USA

[16] NSF Center for Robots and Sensors for Human Well-Being, Purdue University, 156 Knoy Hall, Purdue Polytechnic, West Lafayette, IN 47907, USA

* Correspondence: dvanega@clemson.edu; Tel.: (+1)864-656-1001

Abstract: In this manuscript, we discuss relevant socioeconomic factors for developing and implementing Sensor Analytic Point Solutions (SNAPS) as point-of-care tools to serve impoverished communities. The distinct economic, environmental, cultural, and ethical paradigms that affect economically disadvantaged users, adds complexity to the process of technology development and deployment beyond the science and engineering issues. We begin by contextualizing the environmental burden of disease in select low-income regions around the world, including environmental hazards at work, home, and broader community environment, where SNAPS may be helpful in prevention and mitigation of human exposure to harmful biological vectors and chemical agents. We offer examples of SNAPS designed for economically disadvantaged users, specifically for supporting decision-making in cases of tuberculosis (TB) infection and mercury exposure. We follow-up by discussing the economic challenges involved in phased implementation of diagnostic tools in low-income markets and describe a micropayment-based systems-as-a-service approach (PAPPU), which may be

49 catalytic for adoption of low-end, low-margin, low-research and development SNAPS. Finally,
50 we provide some insights into the social and ethical considerations for the assimilation of
51 SNAPS to improve health outcomes in marginalized communities.

54
55
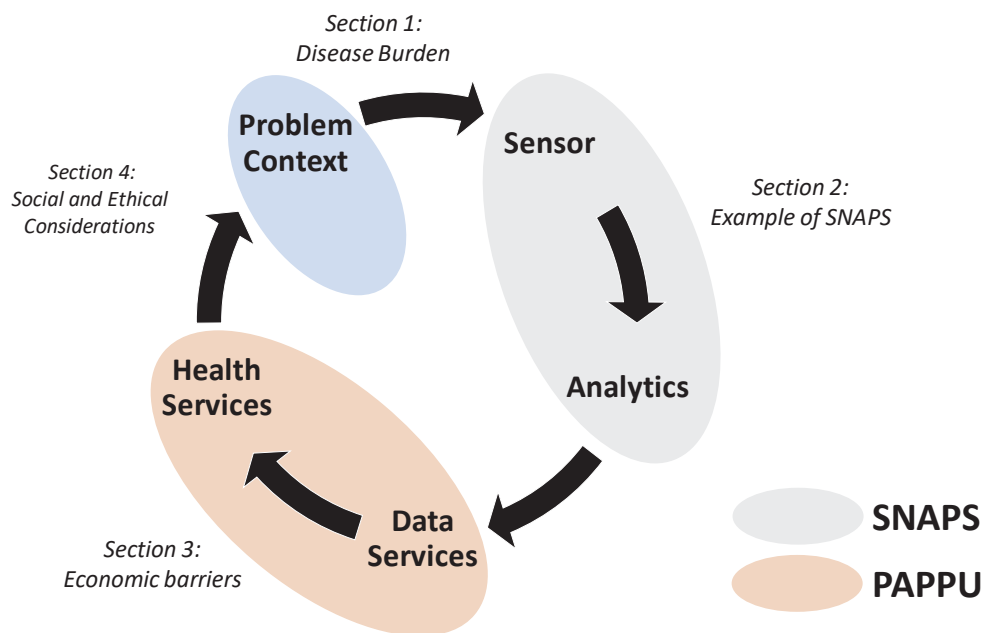
56
57 **Graphical Abstract.** We suggest a framework for using SNAPS (sensor-analytic point solutions)
58 in combination with the financial instrument PAPPU (Pay-A-Penny-Per-Use) for providing
59 health care services for low-income communities. Distinct economic, environmental, and
60 cultural, issues must be taken into consideration. In this review we summarize the environmental
61 burden of disease (section 1), then provide two examples of SNAPS that are at various stages of
62 maturity (section 3) and discuss PAPPU as a potential mechanism to overcome economic barriers
63 (section 4). We close the review by discussing the opportunities and challenges of this approach
64 through the lens of social acceptance.

65 **1. Environmental Burden of Disease**

66 According to the World Health organization, environmental factors including unsafe water,
67 poor sanitation, air pollution, and unintentional exposure to hazardous chemical and biological
68 agents, are root causes for the burden of disease, disability, and death in the developing world
69 [1, 2] Impoverished communities living in polluted and crowded environments are much more
70 susceptible to the double burden of infective and non-communicable diseases, and this situation
71 is often compounded by the lack of adequate infrastructure, weak environmental policy, and
72 deficient or inequitable healthcare systems that disfavor economically challenged users [3-8].
73 Despite the global efforts to reduce poverty, indicators of health disparities between
74 disadvantaged and affluent populations continue to persist. For instance, the 2018 World Bank
75 estimates show that on average, there is a 12-fold difference in mortality rate of infants between
76 low- and high- income populations [9], but in countries experiencing extreme deprivation such
77 as Somalia and Sierra Leone, this rate is nearly 20-fold higher than the average rate in wealthy
78 nations. In 2016, diarrheal diseases linked to poor sanitation and consumption of contaminated

79   food and water were responsible for 1.6 million deaths, 90% of which occurred in South Asia and
80   sub-Saharan Africa [10, 11] The per capita burden of disease from inhalation exposure to airborne
81   polycyclic aromatic hydrocarbons (by-products of fuel combustion) is nearly 33-fold higher in
82   India compared to the USA [12, 13].

83   Nonetheless, it is important to note that due to the myriad ways in which socioeconomic and
84   environmental factors interact, it is very difficult to establish highly detailed associations of single
85   environmental risk factors with epidemiological outcomes [14-17]. Moreover, environmental
86   factors rarely occur in isolation; for example, a population can be exposed to a combination of
87   pollutants from different sources, which could result in additive or synergistic effects and
88   symptoms, making medical diagnostic processes extremely cumbersome (Briggs, 2003). In
89   addition to the limited access to healthcare systems, the problem is compounded by the relatively
90   high cost of clinical testing and may cause many illnesses to go under-reported or mis-diagnosed
91   [18] in economically challenged populations. Despite the complexities involved in linking
92   environmental and socioeconomic factors to epidemiological outcomes, there is no question that
93   such factors can result in serious public health problems, particularly in low-income communities
94   which bear the largest proportion of the burden of environmentally related diseases [19, 20].

95   Undoubtedly, much of the economic strain from both infectious and non-communicable
96   diseases associated with unhealthy environments could be effectively diminished through
97   preventive strategies that tackle associated risk factors [18, 21]. One promising approach for
98   addressing health risk factors in low-income communities is the deployment of integrated
99   technologies for data-informed decision support such as Sensor Analytic Point Solutions
100 (SNAPS). The concept of SNAPS has been recently introduced as part of a platform approach to
101 converge sensor data and analytics to deliver data-informed decision support for a number of
102 applications, including healthcare [21]. Even though thousands of sensors and point-of-care
103 diagnostic tools have been developed in research labs around the world during the past decades,
104 the large majority of these technologies have not yet translated into implementable solutions due
105 to different obstacles including unsuitability of operation under real-world conditions, high
106 fabrication and operation costs (which limits market penetration and profitability), and a lack of
107 convergence with other technologies to yield actionable information for the user [22].

108 Consider for instance the case of diarrheal diseases associated with *E. coli* infection from
109 ingestion of contaminated food or water, which contributes significantly to mortality and
110 morbidity of children under 5 years of age in African and Eastern Mediterranean countries [23].
111 By conducting a literature search on the Web of Science, we found that in the past 10 years 303
112 research articles have been published in peer-reviewed journals portraying the development of
113 *E. coli* biosensors. However, only a small fraction of these papers includes claims such as real-
114 sample testing (~29%), low-cost fabrication (~10%), portability (~9%), and user-friendly operation
115 (~2%) (the complete report from this search is available in the Supplemental Section S1).

116 In this manuscript, we provide examples of SNAPS that have been tested in field conditions,
117 within the context of low-income communities. The first example was developed for assisting
118 early diagnosis of infectious disease and prevention of public health outbreaks, and the second
119 example supports decision making in cases of human exposure to an environmental pollutant.
120 We also propose the concept of Pay-A-Penny-Per-Use (PAPPU) as a potential paradigm to reduce
121 economic barriers to implement SNAPS in economically-deprived regions. The two examples of
122 field-tested SNAPS are at different stages of maturity, which provides insight into the design
123 process and logic flow. Finally, we provide some insights on the social and ethical considerations
124 for the effective use of SNAPS in assisting the users and improving health outcomes in
125 underserved communities.

2. Examples of SNAPS-ART

Near real-time qualitative decisions are often key for rapid response. SNAPS is a tool that uses sensor data to provide response at the point of use with minimal analytics. If two or more factors must be considered by the human-in-the-loop to take a decision, artificial reasoning tools (ART) is implemented. ART is a data fusion layer that combines sensor data and display suggestions or information, on the user's mobile device. In principle, SNAPS is designed to offer "point solutions", which implies a rapid binary output (yes/no) based on the data captured from the sensor signal (for example, sensor binds to an analyte). However, even in rudimentary scenarios, a single source of binary data may fail to provide basic information. Hence, the need for artificial reasoning tools (ART) which are light-weight middleware (software which sits in the "middle") embedded with preliminary logic to decide what is the *meaning* of the data and what information may be conveyed (displayed) for the end-user. By introducing a modular ART tool, the user takes advantage of combinatorial variant configuration menu to change, adapt or introduce new reasoning/logic in the middleware by re-programming the logic "buckets" simply by re-shuffling and inserting the user's preferred choices from a repertoire of pre-programmed logic.

There are many complex layers to system-level solution to ease the environmental burden on impoverished communities. Velez-Torres et al [24] recently developed a circular system framework for integrating analytic tools (such as SNAPS) with social action research (CLISAR). The CLISAR framework is a transdisciplinary approach that involves analytical tools such as sensors for informing community action related to public health, environmental issues, or food security, for example. Beyond simple commercial colorimetric detection strips used in development of CLISAR, information derived from SNAPS can transform this system by supporting decision-making processes aimed at improving the health outcomes of marginalized communities.

Herein, we suggest a conceptual approach for selecting and implementing the type of diagnostic tools for implementation of SNAPS (see **Fig 1**). The examples that follow in the subsequent section used a five-step process that followed a closed-loop approach similar to CLISAR and other circular economic models [25]. The first step is to understand the specific problem as well as the social and economic context where decision-support technology may be needed. The next step is to identify readily available resources and design diagnostic tools for creating a technology portfolio (sensors, analytics software, portable hardware, etc.). The third step involves selection of the most appropriate tools to create SNAPS based on technical capabilities as well as interactive feedback from stakeholders. In step four scientists and end-users test technology prototypes in field conditions using established participatory methodologies. Finally, results from proof-of-concept testing should be used to evaluate and refine the technology. Below, we present two examples of how this conceptual model is applied in real-world settings. The first example is in advanced stage field-testing (refinement and technology improvement, with some elements in the second circular phase), while the second example is in early phase of development (tool selection and technology transfer).
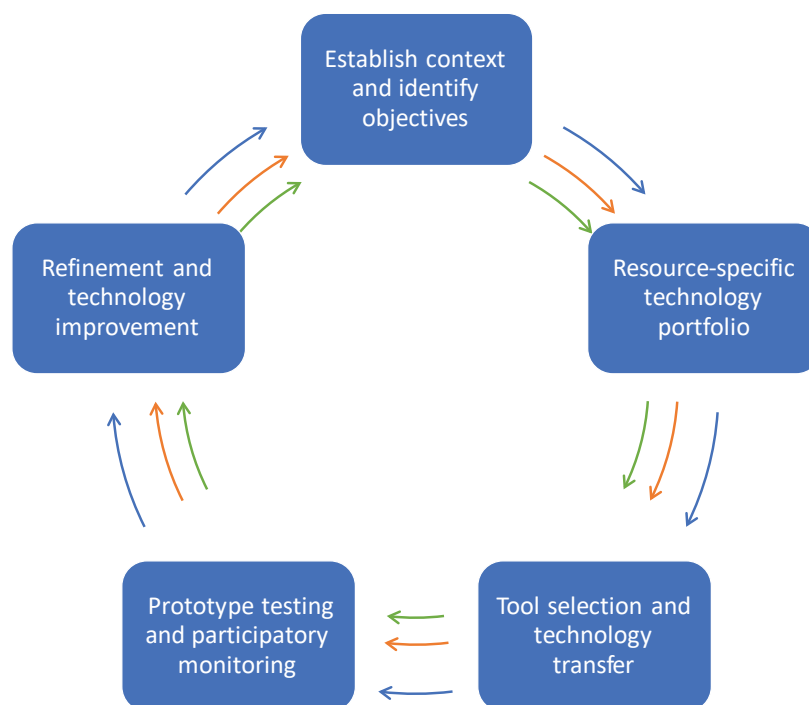
166
167 **Figure 1**. Overview of process in development of SNAPS for the examples shown below. The
168 blue, orange, and green arrows indicate technology evolution using established principles of
169 circular feedback systems. The blue, orange, and green arrows indicate convergence toward a
170 systems-level solution through feedback/refinement pathways.

171 **2.1 *Early assessment of tuberculosis in vulnerable populations***

172 In 2017, 1.6 million people died from tuberculosis (TB) globally and there were 10 million
173 new TB cases that occurred in the same year [26]. TB has surpassed HIV as the leading infectious
174 disease killer worldwide since 2014 [27]. Furthermore, multidrug-resistant and extensively drug-
175 resistant TB (MDR/XDR-TB) are current global public health threats. The 2017 Moscow
176 Ministerial Declaration on ending TB, involving 120 countries and over 800 partners, identified
177 "to advance research and development of new tools to diagnose, treat and prevent TB" as one of
178 four action items [28]. This meeting was followed in 2018 by a United Nations (UN) General
179 Assembly first-ever high-level meeting to accelerate efforts on ending TB [29].

180 Care of TB patients starts with accessible and affordable diagnosis. The majority of TB
181 patients live in poor conditions and in geographically remote areas. Culture-based techniques are
182 the gold standard for diagnosis, but this is relatively expensive and results take 6-8 weeks [30].
183 For decades, TB diagnosis has relied on direct sputum smear microscopy (SSM) in many countries
184 [31]. SSM is fast, inexpensive, facile, and specific for detecting Mycobacterium tuberculosis (Mtb)
185 in high incidence areas [30, 32, 33]. It does not require complex laboratory equipment and is
186 therefore very suitable for low-resource settings especially in populations with varying socio-
187 economic situations [30, 32]. However, SSM's sensitivity is only about 25%-65% compared to
188 culture, with a detection limit of about 10,000 colony forming units per milliliter (CFU/mL) [33,
189 34]. A study comparing culture, SSM, and Xpert MTB/RIF system involving hundreds of
190 specimens showed that SSM had 54% sensitivity for respiratory samples and 50% for non-
191 respiratory samples [35]. Furthermore, smear sensitivity varies with the type of lesion, type and
192 number of specimens, mycobacterial species, staining technique, and the alertness and
193 persistence of the microscopist [34]. In a 2014 survey, 22 high-burden countries (HBCs) conducted
194 77.6 million sputum smears valued at 137 million USD in 42,827 microscopy centers [36]. Of these,

195 61% were performed in the BRICS countries (Brazil, Russian Federation, India, China and South
196 Africa) [36]. On average, 79% of the smears were performed for initial diagnosis in these
197 countries. When converted to 2012 USD, the unit cost for a smear, including materials, labor, and
198 overhead expenses, was 1.77 USD [34]. Studies have shown that the sensitivity of SSM improved
199 significantly when specimens were subjected to liquefaction followed by the concentration of the
200 mycobacteria by overnight sedimentation or centrifugation [33, 37-41]. However, the increased
201 sensitivity provided by these processing methods may not be sufficient to offset their increased
202 cost, complexity, and potential biohazards.

203 Recently, several methods have been developed for the diagnosis and concentration of TB
204 and multi-drug resistant tuberculosis (MDR-TB), such as Xpert MTB/RIF, TB beads, liquid
205 culture, centrifugation, filtration, and line probe assays [42-46]. However, they are not necessarily
206 accessible or affordable for those who need them the most [47]. For example, the World Health
207 Organization recommended the Xpert MTB/RIF in 2010 to diagnose all persons with signs and
208 symptoms of TB. In many studies, the Xpert system was shown to have a sensitivity of 96.8% and
209 a specificity of 99.3% compared to culture as the reference standard [35]. However, if the Xpert
210 MTB/RIF assay (cartridge price of US$9.98) were to be used for all people with presumed TB, the
211 cost would exceed 80% of the total TB spending in countries such as India, Bangladesh, Indonesia
212 and Pakistan [48]. In 2014 and 2015, there were 32.6 and 9.1 SSM for every Xpert MTB/RIF test
213 procured [49]. While these new diagnostic methods are more sensitive and/or specific than SSM,
214 they are oftentimes prohibitively expensive and not easily accessible for those living in low-
215 resource countries where Mtb has a high prevalence.

216 An important aspect of TB is the substantial financial burden placed on patients and their
217 families, not only for treatment costs but also associated costs, such as TB patients who are
218 required to take a leave of absence from work leading to the risk of impoverishment [50].
219 Tanimura et al. reported that on average, 20% of the total cost was due to direct medical costs,
220 20% to direct non-medical costs, and 60% to income loss [51]. On average, the total cost was
221 equivalent to 58% of reported annual individual income and 39% of reported household income
222 [51]. Cost as percentage of income was particularly high among poor people and those with
223 multidrug-resistant TB [51].

224 Accurate, rapid, and cost-effective diagnostic tests are crucial to reducing TB's unacceptably
225 high infection and mortality rates especially for a treatable disease [52]. The ambitious goal of the
226 global "End TB Strategy" to reduce TB incidence by 90% and reduce TB mortality by 95% by the
227 year 2035 will not be achieved without new tools to fight TB [53]. These tools include improved
228 point-of-care diagnostic tests that are delivered to low-income communities and at the first point-
229 of-contact by patients in the healthcare system. These tests should be performed on an easily
230 accessible sample and results be provided in a timely manner, allowing for a quick turnaround
231 time for treatment in a single clinical encounter, hence avoiding loss of patient follow up [53].

232 Thus, our strategy was to develop low-cost biosensing assay for rapid TB detection by
233 employing modern advances in nanoparticle science and glyco-chemistry resulting in sensitivity
234 matching the performance of Xpert MTB/RIF and standard culture. The nanoparticle-based
235 colorimetric biosensing assay (NCBA) is based on the concept of magnetically activated cell
236 enrichment (MACE) technique. In this technique the Mtb cells are isolated and enriched by
237 applying a magnetic field to activate nanoparticle-bound Mtb cells, without using any expensive
238 antibodies and energy-consuming centrifuge instruments and eliminating the need for time-
239 consuming growth of Mtb. The novelty of NCBA includes the utilization of iron oxide
240 nanoparticles with superparamagnetic properties. The use of magnetic nanoparticles (MNPs)
241 offers major advantages due to their unique size and physicochemical properties. The MNP
242 solution is colloidal in nature and it provides stability, which gives rise to both steric and

243 coulombic repulsions. Their nanoscale size results in their higher effective surface areas, lower
244 sedimentation rates, and minimal precipitation due to gravitation forces. The MNPs are coated
245 with glycan to facilitate attachment on the bacterial cell wall through carbohydrate-binding
246 protein sites, providing specificity to the biosensing mechanism.

247 NCBA was validated in 500 sputum samples in Nepal and Mexico [54, 55]. In Nepal, 500
248 sputum samples were tested for TB using SSM, Xpert MTB/RIF, and NCBA. Results showed that
249 for the SSM test, 32 were positive (32+) and 468 were negative (468−); for the Xpert test, 80 were
250 positive (80+) and 420 were negative (420−); and, for the NCBA test, 80 were positive (80+) and
251 420 were negative (420−). The 32+ SSM samples were all positive in Xpert and NCBA. Of the 468-
252 (negative) SSM samples, 48 were positive in both Xpert and NCBA. Table 1 presents the results
253 from SSM test (not shaded) and NCBA test (shaded) using Xpert MTB/RIF as the standard in
254 defining the number of true TB cases and non-TB cases. Statistical analysis of the diagnostic
255 comparison between SSM and NCBA using Xpert MTB/RIF results as gold standard for true
256 cases, and are presented in Table 2. At 95% confidence interval, the results show that SSM has a
257 sensitivity of only 40% (29%−52%), while NCBA has a sensitivity matching that of the Xpert
258 system (95%−100%). Probably because sputum samples were from suspected TB patients, the
259 specificity, positive predictive value (PPV), and negative predictive value (NPV), for SSM and
260 NCBA are very high close to 100%. The accuracy of SSM is 90% (87%−93%), while the accuracy
261 of NCBA is 100% (99%-100%). Given the sample size and nature of the collected samples, the
262 calculated prevalence for this cohort of patients is 16% (80 out of 500).

Table 1. Results using Xpert MTB/RIF as the gold standard for true tuberculosis (TB)
cases and non-TB cases.

SSM Test	True TB cases	Non-TB cases	NCBA Test	True TB cases	Non-TB cases
Positive test	32	0	Positive test	80	0
Negative test	48	420	Negative test	0	420

263

264 **Table 2.** Comparison of diagnostic performance.

Technique	Xpert MTB/RIF as the gold standard, % (95% CI)				
	Sensitivity	Specificity	PPV	NPV	Accuracy
SSM Test	40 (29–52)	100 (99–100)	100	90 (88–91)	90 (87–93)
NCBA Test	100 (95–100)	100 (99–100)	100	100	100 (99–100)

265 The Xpert MTB/RIF system reports bacterial load as very low, low, medium, and high. These
266 categories were used to estimate the bacterial load in SSM and NCBA by matching the
267 corresponding samples with the Xpert system. Table 3 shows a comparison of the detection limit
268 and dynamic range of detection of the two techniques with respect to the Xpert system. Results
269 from NCBA match well with the results from the Xpert MTB/RIF at all levels. On the other hand,
270 SSM could not detect at very low level, detecting only 14% at low level, 48% at medium level, and
271 79% at high level. Sensitivity of NCBA matches well with the Xpert system at all bacterial loads,
272 while SSM increases linearly with increasing bacterial load ($y = 0.27x − 0.33$, $R2 = 0.97$). TB positive

273 samples are normally distributed around the medium level. SSM has 48% detection at the
274 medium level, which is very close to the calculated sensitivity of 40% (29%-52%) in Table 2.

Table 3. Detection limit and dynamic range of detection of the two techniques with respect to the Xpert MTB/RIF categories.

Xpert MTB/RIF Categories	Very Low	Low	Medium	High	Total
Xpert MTB/RIF	10	22	29	19	80
NCBA	10	22	29	19	80
SSM	0	3	14	15	32
% Detection (NCBA/Xpert)	100%	100%	100%	100%	
% Detection (SSM/Xpert)	0%	14%	48%	79%	

275   Thus, with the Xpert system as the standard, the sensitivity of NCBA is in accordance with
276 the Xpert system (100%), while SSM's sensitivity is only 40%. NCBA could detect very low AFB
277 concentrations at 102 CFU/mL, two orders of magnitude lower than SSM. This nanotechnology-
278 based detection is rapid (10–20 min), inexpensive (0.10 USD/test), simple, and easily scalable
279 (Figure 1). According to Nepal's Ministry of Health, a TB diagnostic test with 70% sensitivity (and
280 treatment cure of 85%) would save 300,000 lives over the next five years [56]. This NCBA
281 technique has a high potential to support and transform the TB control program in Nepal and in
282 other high-prevalence low-resource countries. Implementation in rural areas would help to
283 increase case finding and case notification, and would support programs targeted against drug-
284 resistant TB. Nepal has close to 600 microscopy centers that can support the immediate
285 implementation of this technology in the country. Similarly, it is applicable in many of the high
286 TB-burden countries. This technique can be performed in rural communities and at the first point-
287 of-contact by patients in the healthcare system. Results are obtained in less than 30 min, allowing
288 for a quick turnaround time for treatment in a single clinical encounter. Desikan 2013
289 hypothesized that a universally accessible and rapid detection method with a sensitivity of 85%
290 and specificity of 97% could potentially save 392,000 lives annually worldwide [32]. Thus, the
291 developed NCBA technology may enhance the "End TB Strategy" towards a TB-free world.
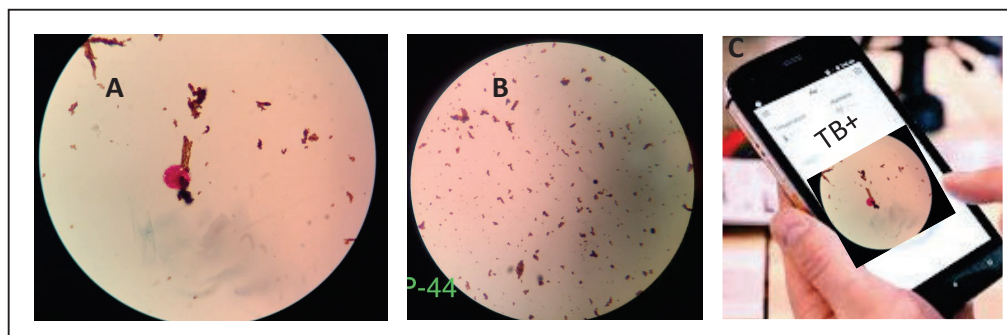
Figure 2. typical NCBA results for TB+ and TB− sputum samples, as viewed through the eyepiece of the bright field microscope. A) TB positive sample (clumped red GMNP-AFB complex surrounded by brown GMNPs). B) TB negative sample (dispersed brown GMNP). C) Schematic of smartphone app for image processing and display of test results.

302 **2.2** *Alerting mercury exposure in artisanal gold mining communities*

303   In South America, Africa, and Asia, millions of individuals are exposed to dangerous levels
304 of mercury concentrations as a result of Artisanal Small-Scale Gold Mining (ASGM) [57]. ASGM

305   is a rudimentary gold mining approach performed by individuals or groups with little or no
306   mechanization, often in informal (illegal) operational settings with toxic chemicals [58]. ASGM is
307   composed of three main steps: crushing the ore into fines, mixing the fines with liquid mercury,
308   and separating the mercury from gold by evaporating the mercury [59]. Often in unregulated
309   occupational conditions, workers perform mercury evaporation using open pit, which not only
310   has severe adverse health effects for the workers that inhale the mercury vapor but also releases
311   the toxic vapor into the environment. ASGM recently exceeded combustion of coal as the leading
312   anthropogenic source for mercury emissions globally [60]. Risk of exposure to mercury can lead
313   to detrimental effects on the nervous, immune, reproductive, and digestive systems, induce
314   infertility, reduce mental function, and kidney failure [61-65].

315   The global responsibility for reducing mercury emissions was recognized by the Minamata
316   Convention in Switzerland in 2013. At the convention, over 140 countries signed a treaty
317   committing to protect human health from mercury exposure [60]. The signatory countries
318   pledged to "ban new mercury mines, phase-out existing mines, ensure the phase out and phase
319   down of mercury use in a number of products and processes, develop control measures for
320   emissions, and regulate the informal sector of ASGM" [60]. In order to mitigate mercury exposure
321   and regulate mining operations, it is prudent for marginalized communities to monitor the
322   presence of mercury in their water through low-cost, rapid, and facile devices.

323   Several analytical methods have been developed for mercury determination in water.
324   Standard laboratory techniques include cold vapor atomic absorption spectroscopy (CV-AAS)
325   [66, 67], cold vapor-atomic fluorescence spectrometry (CV-AFS) [68, 69] and inductively coupled
326   plasma mass spectrometry (ICP-MS) [70, 71]. These spectroscopic techniques are highly sensitive
327   and accurate but are often impractical for environmental applications due to the high cost of
328   analysis. In addition, these standard methods require extensive user training, and the results
329   often require days or even weeks to produce results, making them less suitable for rural
330   communities [72-74]. Some field capable units are commercially available, namely based on direct
331   mercury analysis (DMA) and handheld nanosensors/biosensors [75, 76]. DMA is based on the
332   principle of thermal decomposition (vaporization), followed by amalgamation and subsequent
333   atomic absorption spectroscopy. While extremely accurate, DMA is cost prohibitive for low-
334   income communities since commercial prices of US-manufactured equipment range between
335     $13k and $30k USD. Perhaps inexpensive nanosensors/biosensors coupled with low-cost
336   electrochemical techniques on portable devices are likely to be more suitable as a tool for on-site
337   analysis of mercury, especially where ASGM is in practice.

338   While there are many types of transduction methods for low-cost determination of mercury,
339   electrochemical methods are sensitive and quantitative, and may be the mechanism of choice for
340   cost-effective rapid detection in the field [77]. The most common electrochemical method for ionic
341   mercury detection is anodic linear stripping voltammetry (ASV) techniques [72, 78]. ASV is a two-
342   step method of deposition/accumulation during reduction of mercury ions and stripping during
343   oxidation of mercury ions along the surface of the electrode. As the mass transfer limit is reached
344   in the reaction, the oxidative current forms a well-defined peak, which can be used to calculate
345   the concentration of mercury in the sample [79]. The efficiency of any electrochemical stripping
346   test can be determined by calculating the percent change in oxidative current relative to baseline.

347   Carbon-based nanomaterials are a popular choice for improving electrochemical detection
348   of mercury, as this type of materials exhibit high surface area, strong mechanical strength,
349   excellent thermal conductivity, and high conductivity [80-81]. Some of the carbon nanomaterials
350   in recent literature include glassy carbon [83, 84], carbon nanotubes [85], graphene [86], and
351   reduced graphene oxide [87]. While each of these nanocarbon materials is efficient for mercury
352   detection via stripping voltammetry, some of the materials are complicated to fabricate and

353   exhibit poor water solubility [88]. Among carbon nanomaterials, graphene and reduced graphene
354   oxide (rGO) have the highest water solubility, and one of the lowest fabrication costs. For these
355   reasons, there is a growing trend to develop disposable, low-cost, graphene-based electrodes for
356   field applications.

357   Examples of low-cost graphene electrodes include screen-printed electrodes and conductive
358   paper and plastic [72, 89]. In 2014, Lin et al. (2014) [90] discovered a low-cost, one-step,
359   conductive material when reducing graphene on a commercial polymer with a carbon dioxide
360   infrared laser. Since then, multiple researchers have shown that laser scribing could be used for
361   designing electrodes to sense biomolecules using infrared and ultra-violet light lasers [91-94].
362   While graphene is indeed a useful material in sensing, one of the problems is the tendency of
363   graphene and graphene oxide for binding a host of materials in aqueous phase [95]. For this
364   reason, sensor labs typically metallize graphene electrodes with a noble metal that has a specific
365   interaction with mercury ions. These metals can be deposited using simple electrodeposition
366   methods, or advanced techniques such as pulsed sonoelectrodeposition [96]. Recently,
367   Abdelbasir et al. 2018 [97] showed that copper nanoparticles recovered from waste cables can be
368   used to detect ionic mercury using linear sweep stripping voltammetry (LSSV).

369   Low-cost, portable, mobile phone-based acquisition systems have been developed for
370   mercury analysis in the field [98]. While this is significant for deploying sensors in low-income
371   regions, the inexpensive-portable sensor-systems lack data analytics capability to transform the
372   data into *meaningful information* which could be useful for the user. For example, the maximum
373   concentration level for inorganic mercury in drinking water is 6 ppb [99]. However, bodyweight,
374   ingestion rate, length of exposure, form and pathway of the contaminant, health of the individual,
375   and concentration of mercury influences the degree of mercury toxicity [101-102]. Thus, a SNAPS
376   tool may assist communities in acquiring data and extracting actionable information for decision
377   support.

378   Our group is currently working on developing the SNAPS platform for estimating the
379   toxicity risk associated with the ingestion of mercury-contaminated water. This SNAPS platform
380   is composed of a disposable graphene-nanocopper sensor coupled with a low-cost handheld
381   potentiostat, and a smartphone. The working mechanism of the platform starts with the detection
382   of mercury present in the sample using the graphene-nanocopper sensor. Next, selective
383   electrochemical interactions between mercury and the electrode generates an electrical signal.
384   The electrical signal is acquired and processed by the potentiostat to produce a current output.
385   Then, computer software records the current output and transforms it into concentration data via
386   calibration curves. Finally, a smartphone app is used by the user to enter the data for the
387   following parameters: mercury concentration in water (from the sensor), bodyweight of the user,
388   water ingestion rate, and length of exposure. Based on these parameters, the app runs an
389   algorithm that includes a hazard quotient formula to generate an estimation of the risk of toxicity
390   for the user [103-106].

391   We recently conducted a proof-of-concept demonstration of this SNAPS platform in a rural
392   area that has been dramatically impacted by ASGM, known as La Toma in Cauca, Colombia.
393   Even though this SNAPS platform is in an early stage of development, it represents an example
394   of how rural communities in developing countries may use sensors as a service to access data on
395   mobile devices and extract actionable information to help make informed decisions. Figure 3
396   shows the progression of the proof-of-concept demonstration of the technology.

397   Mercury enters natural aquatic systems primarily due to the burning of mercury amalgam during
398   the extraction of gold from raw ore.

399
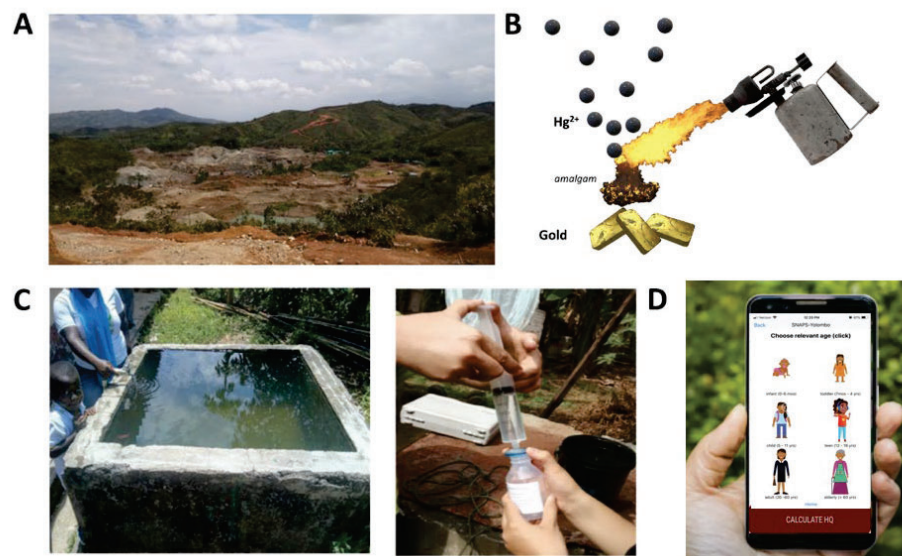
Figure 3. Demonstration of SNAPS tool for assessing risk due to inadvertent consumption of mercury in drinking water for gold mining communities in Colombia. The first step was to (**A**) characterize the local socioeconomic dynamics and (**B**) identify related routes of mercury exposure (in this case from smelting of amalgam). (**C**) Together with community members, we collected samples from local water sources. (**D**) These samples were tested with nanomaterial-enabled sensors. (**D**) Concentration data derived from sensors was transformed into customized information about toxicity risk for specific user groups using a mobile app.

3. Can we overcome the economic barriers for distributing diagnostic tools in low-income settings?

Framing the issue of diagnostic tools in the context of technology leads us to recognize a vast spectrum. On one hand, ideas about telemedicine proposed about 100 years ago [107], and on the other hand, milestones in computational speed from about 100 days ago [108]. It may be justifiable to suggest that technological barriers may not be the primary reason why many diagnostic tools are still absent from communities under economic constraint. The powerful incentive of lucrative profitability, in the short term, may not be realized by serving impoverished regions.

Transaction cost [109] may be the over-arching factor which may have multiple interpretations [110] but appears to be the economic barrier with respect to the reasons why accelerating the rate of diffusion of diagnostic tools in distressed communities continues to pose difficult challenges [111-113]. We must focus on value to the user or the extent of the benefit to the beneficiary's environment and/or ecosystem (for example, early diagnosis of tuberculosis in a patient may save the entire village from infection and epidemic). However, delivery of value is inextricably linked to cost, unless it is aimed to deliver philosophical or mythical messages [114].

In over-simplified terms, the convergence of the cost of the product and the cost to deliver the service contributes to transaction cost [115]. A plethora of costs and cost-incurring processes are involved but we shall bypass the details. The physical product (in this case is the sensor) and the service is the solution delivery (SNAPS). Academics cannot control cost, but their contribution can impact implementation and use. A low-cost sensor from a lab must be manufactured, calibrated, evaluated and sufficiently scaled if the outcome can still be claimed as a "low-cost" sensor, capable of delivering value with respect to maintaining a certain pre-agreed quality of

430 service (QoS) in keeping with the key performance indicators (KPI) that the users desire, demand
431 or deem necessary.

432 In addition, a working sensor delivered to a user is useless without a visualization system to
433 capture the data from the sensor. Stand-alone visualization devices (for example, blood glucose
434 home monitors with dedicated devices to read the blood glucose strip and deliver data readout)
435 will add inordinate cost to the system. The alternative is to use a mobile phone as a platform to
436 visualize the data from the sensor. Signal transduction from the sensor to the mobile phone calls
437 for multiple layers of tools, technology and software (middleware), in addition to the functional
438 use of a mobile phone. The presence of a mobile phone in any environment is contingent upon
439 available cellular and/or wireless infrastructure to support its use. It may not be prudent to
440 assume the presence of telecommunications infrastructure despite the penetration of such
441 services, globally [116-119]. Thus, even if a working sensor is at hand, the obvious process of
442 signal to data transition and visualization of the data involves multiple layers of capital expenses
443 (infrastructure cost) as well as associated technologies and software.

444 Assuming that the above layers are in working order, the sensor data meets a "dead end"
445 upon data visualization. A number (with units) is only meaningful if there is a relevant
446 framework for interpreting such data, for example, the combination of sensor data from mercury
447 contamination expressed in terms of a hazard quotient score, which uses other vital pieces of
448 information to assess health risk. It is the *delivery of information* based on sensor data that drives
449 value. Taken together, the physical product is no longer the focal point of value. Information
450 pertaining to the health of the user is the service which delivers value to the user. Transaction
451 cost, therefore, is no longer a product-based entity but the cost of service which must be feasible
452 for the service to be delivered, disseminated, and adopted by a community.

453 Overcoming the economic barriers to deliver SNAPS will be virtually impossible if the
454 chasm between product and service continues to overshadow the concept of value delivery to the
455 user. The economic principle, which may work in impoverished nations, is rooted in micro-
456 finance and micro-payments with low transaction costs [120, 121]. The paradigm shift from
457 "product sales" to delivery of "service" involves combining the product with resources
458 (including retail mobile banking, infrastructure, telecommunications, cybersecurity, customer
459 service). Users pay only when they use the service. The latter lowers the transaction cost and
460 hence the barrier to entry into vast markets of low-income users. Not the product, but the *user*
461 *experience* is the pivotal fulcrum for the inversion of traditional business models in the era of the
462 internet of things (IoT) [122].

463 The PAPPU model was epitomized by the plain old telephone system (POTS), where the
464 user paid only the "charge per call" which was reasonably affordable even if the per capita
465 income was low. In this paper we advocate for PAPPU as a metaphor for ethical profitability
466 through social business models. In principle, the user may pay a penny for each use of a SNAP
467 (suggested but not restricted to one penny). The "penny" is a placeholder for the financial design
468 of an ultra low-cost nano-payment model, which, in the real world may represent one Rupee
469 (INR), one Yuan (CNY, RMB) or one Peso (COP). The PAPPU metaphor may evolve to become
470 the *generalized* monetization mantra signifying pay-a-price-per-unit wherever the principles of
471 IoT may be deployed or embedded as a digital by design metaphor including ubiquitous sensing.
472 The diffusion of connectivity may serve as a tool and IoT may be catalytic as a platform to better
473 facilitate the practice of equality, equity and égalité. PAPPU offers *an economic instrument* for
474 businesses to build a profit model based on economies of scale to serve low-income communities
475 and abide by ethical profitability. PAPPU offers an alternative strategy for enterprises and
476 businesses who are seeking to engage with the next billion users, albeit profitably, but within the

477   realms of ethical profitability which can be *sustained* by the per capita income of these
478   communities.

479   The concomitant growth of infrastructure (for example, affordable access to low latency,
480   reduced jitter, high bandwidth wireless telecommunications, 5G, trusted mobile banking) may
481   be necessary to pave the road for pursuit of PAPPU. The ability to escape the dead weight of old
482   technology in the developing world may accelerate the implementation of PAPPU as an integral
483   part of the socio-economic fabric of a product-less, service-based economy where payment per
484   unit of service (1 liter of municipal water, 1 kilo-watt hour of energy, 1 gallon of sanitation waste)
485   may become the new normal.

486   Implementing PAPPU may require alliances or public-private partnerships or global
487   consortia with an altruistic fervor to pay and pave for the synergistic integration necessary, to
488   promote SNAPS as a service in low-income communities. The challenge is to bring to the table
489   global organizations, benevolent individuals and thoughtful governments, who may choose to
490   lead this effort to channel science to serve society, for the less fortunate. We need new eyes,
491   unbridled imagination and moral fabric of synergistic solutions that can wrap around, not to
492   isolate, but to protect, provide and promote acceptable solutions for remediable injustices.

493   **4.Social and ethical considerations for the development and implementation of SNAPS**

494   Cultural and ethical considerations are inextricably linked with the transformation of SNAPS
495   from an academic vision to real-world implementations which may actually help people.
496   Academics must remain cognizant of their ethical responsibility to discourage misapplication and
497   dissemination of misinformation about their inventions. In this section, we attempt to analyze
498   some potential interactions between the social and technological domains, as well as how
499   democratic approaches for technology creation and diffusion could favor the improvement of
500   health outcomes for disadvantaged communities.

501   Since the introduction of the technology acceptance model (TAM) decades ago, several extended
502   versions of this archetype have been proposed to elaborate a more comprehensive framework for
503   predicting people's intention to use a particular product or service [123-125]. TAM and its
504   variants have served as the guiding rationale behind R&D for a variety of commercial
505   technologies that are mass-produced, including healthcare devices [126]. However, it may be
506   inadequate in the context of technology development for low-income communities [127]. It is
507   worth noting that the ultimate goal of TAM and related models is to forecast user behavior across
508   a broad range of consumer populations, which means that the model focuses on highly generic
509   predictors of technology acceptance. For instance, TAM does not explicitly include any cultural
510   or social variables, which is a significant limitation since social constructs may contribute
511   significantly to the variance in users' attitudes towards technology [125, 128]. However, the goal
512   of the SNAPS with PAPPU concept is to provide an affordable sensor-analytics service platform
513   to support decision-making and enhancement of health outcomes for economically challenged
514   groups. Thus, a useful model to guide the development of SNAPS should include bi-directional
515   communication between researchers and users, and perhaps motivate researchers and users to
516   change or adapt or better inform their behavior [129].

517   Trust in the technology [130] is quintessential for adoption and continued use because
518   technology could be equally euphemistic for a double-edged sword [131, 132]. Driving positive
519   impacts from the introduction of SNAPS in low-income regions may involve not only the
520   transfer of fully functional technology but also the empowerment of the beneficiary
521   communities by enabling local mastery of the technology along with the possibility to
522   reproduce and even adapt the technology to the local conditions. We believe this open-source

523 approach to technology adoption is auspicious for supporting marginalized communities,
524 especially when trying to avoid the known failures of the charitable approach of technology
525 leapfrogging. For example, the WHO estimates that only 10-30% of the medical devices donated
526 to developing countries are used as intended; the remaining 70-90% end-up being dumped in
527 landfills, thus contributing to more pollution problems and environmental health risks [133].
528 This situation is explained not only by the incompatibility of the technology with the locally
529 available infrastructure, but also to the lack of local capacity to adapt or fix the donated devices
530 once they break [134]. Additionally, dependence on foreign technologies could lead to an
531 imbalance of power in which the users have no option other than relying on the willingness of
532 external entities to continue to deliver much-needed technology in their regions. Thus, if the
533 goal is to make technology work effectively on behalf of society, we must divert from the
534 mainstream *handed-down from the top* approach and enable society to create and transform
535 technology in meaningful ways, in dispersed regions, and from the bottom-up.

536 Engaging the community through operational transparency may prevent public anxiety and
537 may also facilitate proper implementation of the technology. Users' understanding of the
538 limitations and potential risks associated with SNAPS could be vital for setting clear expectations
539 about SNAPS-assisted testing while avoiding misapplications of the technology. As Wallace et
540 al. point out, misuse of many direct-to-consumer screening tests could have caused an
541 unnecessary increase in healthcare costs due to people's overreaction to inaccurate readings from
542 direct-to-consumer screening tests, and their subsequent demand for further testing with
543 advanced clinical technology [130]. However, this concern is mostly relevant for developed
544 countries in which people have access to healthcare systems where clinical testing is readily
545 available for patients. In low-income settings, such as remote rural areas in developing countries,
546 health care services are often dysfunctional or completely inaccessible. For marginalized
547 communities, information from SNAPS could instead drive actions aimed at limiting the
548 exposure to harmful biological vectors and chemical agents. Thus, communities living in
549 territories that suffer from prolonged government abandonment could greatly benefit from the
550 democratic adoption of SNAPS to make informed decisions and solve their problems with more
551 autonomy. Nonetheless, we agree that transparency and accountability in technology
552 deployment are paramount for protecting the users' rights and integrity.

553 **5. Conclusions**

554 Monitoring environmental contamination is essential to protect the public from diseases and
555 other health issues. This monitoring requires accurate and accessible detection technologies to
556 ensure quality control and early warning capabilities for users to minimize negative impacts
557 (Figure 4). The framework of SNAPS with PAPUU have the potential to pave the way for
558 economically viable systems that can potentially be applied as tools for reducing local
559 environmental risks and mitigate health problems that derive from them. We envision that the
560 use of SNAPS will increase low-income communities' participation in the public/government
561 planning process by providing data they can use to fight for their right to public health care,
562 clean water and adequate sanitation. By bridging smart technology with basic needs and public
563 health, SNAPS will advance our understanding of how information can change public
564 participation, having low-income communities' representatives as 'change agents' that
565 influence public policies and planning. These communities' representatives benefit from rights-
566 based arguments, evidence-based research, and effective data analyses. SNAPS have the
567 potential to serve as an illustration of how empowering impoverished communities in their
568 local context, can strengthen democratic practice in their region. Grounded on an integrated
569 perspective that takes into account cultural and ethical considerations, we foresee that SNAPS
570 will shed some light to improve implementation of public health plans in underserved
571 communities by increasing public participation in planning. Moreover, SNAPS could

572   potentially become a new approach to achieve the United Nations Sustainable Development
573   Goals 3 and 6: ensure healthy lives and promoting the well-being at all ages and ensure access
574   to water and sanitation for all, respectively. Furthermore, it would also help empower
575   impoverished communities to obtain the rights they have been promised such as basic
576   sanitation, clean water, and adequate health care services.

577

578

579   **Figure 4.** SNAPS converges with PAPPU to establish a framework for sensor-as-a-service. The
580   paradigm is rooted in economic, ethical, cultural, and environmental core values that
581   synergistically act as a catalyst for the democratization of healthcare in underserved
582   communities,

594   **References**

595

1. Briggs, D., Environmental pollution and the global burden of disease. *British medical bulletin* **2003**, *Volume 68*, pp. 1-24. 10.1093/bmb/ldg019.

2. World Health Organization and United Nations Environment Programme, Health and environment: tools for effective decision-making: review of initial findings. 2004.

3. Cureton, S., Environmental victims: environmental injustice issues that threaten the health of children living in poverty. *Reviews on environmental health* **2011**, *Volume 26*, pp. 141-147.

4. Boutayeb, A., The double burden of communicable and non-communicable diseases in developing countries. *Transactions of The Royal Society of Tropical Medicine and Hygiene* **2006**, *Volume 100*, pp. 191-199. 10.1016/j.trstmh.2005.07.021 %J Transactions of The Royal Society of Tropical Medicine and Hygiene.

5. Gwatkin, D.R.; Bhuiya, A.; Victora, C.G., Making health systems more equitable. *The Lancet* **2004**, *Volume 364*, pp. 1273-1280.

6. Marmot, M., Social determinants of health inequalities. *The lancet* **2005**, *Volume 365*, pp. 1099-1104.

7. Evans, G.W.; Kantrowitz, E., Socioeconomic status and health: the potential role of environmental risk exposure. *Annual review of public health* **2002**, *Volume 23*, pp. 303-331.

8. Waller, L.A.; Louis, T.A.; Carlin, B.P., Environmental justice and statistical summaries of differences in exposure distributions. *Journal of Exposure Science Environmental Epidemiology* **1999**, *Volume 9*, pp. 56.

9. Mortality rate, infant (per 1,000 live births). Available online: https://data.worldbank.org/indicator/SP.DYN.IMRT.IN?end=2018&start=2018&type=points& view=bar (accessed on September 27 2019).

10. Moraga, P.; Collaborators, G.C.o.D., Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **2017**, *Volume 390*, pp. 1151-1210.

11. Collaborators, G.D.D., Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis* **2018**, *Volume 18*, pp. 1211-1228.

12. Etchie, A.T.; Etchie, T.O.; Shen, H.; Pillarisetti, A.; Popovicheva, O., Burden of disease at the same limit of exposure to airborne polycyclic aromatic hydrocarbons varies significantly across countries depending on the gap in longevity. *Ecotoxicology environmental safety* **2019**, *Volume 180*, pp. 420-429.

13. Schraufnagel, D.E.; Balmes, J.R.; Cowl, C.T.; De Matteis, S.; Jung, S.-H.; Mortimer, K.; Perez-Padilla, R.; Rice, M.B.; Riojas-Rodriguez, H.; Sood, A., Air pollution and noncommunicable diseases: a review by the Forum of International Respiratory Societies' Environmental Committee, Part 2: air pollution and organ systems. *Chest* **2019**, *Volume 155*, pp. 417-426.

14. Valdivia-Rivera, S.; Martínez-Cano, A.; Aguirre-García, G.; Lizardi-Jiménez, M., Hydrocarbon water-pollution related to chronic kidney disease in Tierra Blanca, a perfect storm. *Environment international* **2018**, *Volume 121*, pp. 1204-1209.

15. Kumar, V.; Parihar, R.D.; Sharma, A.; Bakshi, P.; Sidhu, G.P.S.; Bali, A.S.; Karaouzas, I.; Bhardwaj, R.; Thukral, A.K.; Gyasi-Agyei, Y., Global evaluation of heavy metal content in surface water bodies: a meta-analysis using heavy metal pollution indices and multivariate statistical analyses. *Chemosphere* **2019**, *Volume* pp. 124364.

16. Johannson, K.A.; Balmes, J.R.; Collard, H.R., Air pollution exposure: a novel environmental risk factor for interstitial lung disease? *Chest* **2015**, *Volume 147*, pp. 1161-1167.

17. Hu, C.-Y.; Gao, X.; Fang, Y.; Jiang, W.; Huang, K.; Hua, X.-G.; Yang, X.-J.; Chen, H.-B.; Jiang, Z.-X.; Zhang, X.-J., Human epidemiological evidence about the association between air pollution exposure and gestational diabetes mellitus: Systematic review and meta-analysis. *Environmental research* **2019**, *Volume* pp. 108843.

18. Stanaway, W.D.; GBD 2017 Risk Factor Collaborators, Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **2018**, *Volume 392*, pp. 1923-1994.

19. Prüss-Ustün, A.; Wolf, J.; Bartram, J.; Clasen, T.; Cumming, O.; Freeman, M.C.; Gordon, B.; Hunter, P.R.; Medlicott, K.; Johnston, R., Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low-

and middle-income countries. *International journal of hygiene environmental health* **2019**, *Volume 222*, pp. 765-777.

20. Ostro, B.; Spadaro, J.V.; Gumy, S.; Mudu, P.; Awe, Y.; Forastiere, F.; Peters, A., Assessing the recent estimates of the global burden of disease for ambient air pollution: methodological changes and implications for low-and middle-income countries. *Environmental research* **2018**, *Volume 166*, pp. 713-725.

21. Bauer, U.E.; Briss, P.A.; Goodman, R.A.; Bowman, B.A., Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *The Lancet* **2014**, *Volume 384*, pp. 45-52.

22. From the lab to real-world use. *Nature Sustainability volume* **2019**, *Volume 2*, pp. 989. https://doi.org/10.1038/s41893-019-0435-7.

23. Anderson IV, J.D.; Bagamian, K.H.; Muhib, F.; Amaya, M.P.; Laytner, L.A.; Wierzba, T.; Rheingans, R., Burden of enterotoxigenic Escherichia coli and shigella non-fatal diarrhoeal infections in 79 low-income and lower middle-income countries: a modelling analysis. *The Lancet Global Health* **2019**, *Volume 7*, pp. e321-e330. https://doi.org/10.1016/S2214-109X(18)30483-2.

24. Vélez-Torres, I.; Vanegas, D.C.; McLamore, E.S.; Hurtado, D., Mercury pollution and artisanal gold mining in Alto Cauca, Colombia: woman's perception of health and environmental impacts. *The Journal of Environment Development* **2018**, *Volume 27*, pp. 415-444. https://doi.org/10.1177/1070496518794796.

25. World Health Organization (WHO), Circular Economy and Health: Opportunities and Risk. 1.52. 2018.

26. Tuberculosis (TB). Available online: https://www.who.int/news-room/fact-sheets/detail/tuberculosis (accessed on November 13 2019).

27. Yang, T.; Zhong, J.; Zhang, J.; Li, C.; Yu, X.; Xiao, J.; Jia, X.; Ding, N.; Ma, G.; Wang, G., Pan-genomic study of Mycobacterium tuberculosis reflecting the primary/secondary genes, generality/individuality, and the interconversion through copy number variations. *Frontiers in microbiology* **2018**, *Volume 9*, pp. 1886. https://doi.org/10.3389/fmicb.2018.01886.

28. Dias, H.M.Y.; Pai, M.; Raviglione, M.C., Ending tuberculosis in India: A political challenge & an opportunity. *The Indian journal of medical research* **2018**, *Volume 147*, pp. 217. https://doi.org/10.4103/ijmr.IJMR_660_18.

29. WHO | UN General Assembly High-Level Meeting on ending TB. Available online: http://www.who.int/tb/features_archive/UNGA_HLM_ending_TB/en/ (accessed on November 15 2019).

30. Olaru, I.D.; Heyckendorf, J.; Grossmann, S.; Lange, C., Time to culture positivity and sputum smear microscopy during tuberculosis therapy. *PloS one* **2014**, *Volume 9*, pp. e106075. https://doi.org/10.1371/journal.pone.0106075.

31. Hobby, G.L.; Holman, A.P.; Iseman, M.D.; Jones, J.M., Enumeration of tubercle bacilli in sputum of patients with pulmonary tuberculosis. *Antimicrobial agents chemotherapy* **1973**, *Volume 4*, pp. 94-104.

32. Desikan, P., Sputum smear microscopy in tuberculosis: is it still relevant? *The Indian journal of medical research* **2013**, *Volume 137*, pp. 442.

33. Steingart, K.R.; Ng, V.; Henry, M.; Hopewell, P.C.; Ramsay, A.; Cunningham, J.; Urbanczik, R.; Perkins, M.D.; Aziz, M.A.; Pai, M., Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *The Lancet infectious diseases* **2006**, *Volume 6*, pp. 664-674. https://doi.org/10.1016/S1473-3099(06)70602-8.

34. Demers, A.-M.; Verver, S.; Boulle, A.; Warren, R.; Van Helden, P.; Behr, M.A.; Coetzee, D., High yield of culture-based diagnosis in a TB-endemic setting. *BMC infectious diseases* **2012**, *Volume 12*, pp. 218. https://doi.org/10.1186/1471-2334-12-218.

35. Afsar, I.; Gunes, M.; Er, H.; Sener, A.G., Comparison of culture, microscopic smear and molecular methods in diagnosis of tuberculosis. *Revista Española de Quimioterapia* **2018**, *Volume 31*, pp. 435.

36. Kik, S.V.; Denkinger, C.M.; Chedore, P.; Pai, M., Replacing smear microscopy for the diagnosis of tuberculosis: what is the market potential? *European Respiratory Journal* **2014**, *Volume 43*, pp. 1793-1796. https://doi.org/10.1183/09031936.00217313.

37. Allen, V.; Nicol, M.; Tow, L., Sputum processing prior to Mycobacterium tuberculosis detection by culture or nucleic acid amplification testing: a narrative review. *Res Rev J Microbiol Biotechnol* **2016**, *Volume 5*, pp. 96-108.

38. Das, S.; Narang, P.; Nagamiah, S.; Mishra, P.; Deotale, V.; Mendiratta, D., Evaluation of variants of carbol fuchsin solution to stain acid-fast bacilli in-situ by the pot method. *The International Journal of Tuberculosis Lung Disease* **2015**, *Volume 19*, pp. 1470-1475. https://doi.org/info:doi/10.5588/ijtld.15.0272.

39. Ho, J.; Marks, G.; Fox, G., The impact of sputum quality on tuberculosis diagnosis: a systematic review. *The International Journal of Tuberculosis Lung Disease* **2015**, *Volume 19*, pp. 537-544. https://doi.org/10.5588/ijtld.14.0798.

40. Munir, M.; Shabbir, I.; Iqbal, R.; Khan, S.U., Comparison of detection of acid fast bacilli in clinical samples by AFB smear microscopy and culture in the diagnosis of tuberculosis in a tertiary care setting. *Pakistan Journal of Chest Medicine* **2015**, *Volume 15*, pp.

41. Shea, Y.R.; Davis, J.L.; Huang, L.; Kovacs, J.A.; Masur, H.; Mulindwa, F.; Opus, S.; Chow, Y.; Murray, P.R., High sensitivity and specificity of acid-fast microscopy for diagnosis of pulmonary tuberculosis in an African population with a high prevalence of human immunodeficiency virus. *Journal of clinical microbiology* **2009**, *Volume 47*, pp. 1553-1555. https://doi.org/10.1128/JCM.00348-09.

42. Cudahy, P.; Shenoi, S.V., Diagnostics for pulmonary tuberculosis. *Postgraduate medical journal* **2016**, *Volume 92*, pp. 187-193. https://doi.org/10.1136/postgradmedj-2015-133278.

43. Dunn, J.J.; Starke, J.R.; Revell, P.A., Laboratory diagnosis of Mycobacterium tuberculosis infection and disease in children. *Journal of clinical microbiology* **2016**, *Volume 54*, pp. 1434-1441. https://doi.org/10.1128/JCM.03043-15.

44. Nyendak, M.R.; Lewinsohn, D.A.; Lewinsohn, D.M., New diagnostic methods for tuberculosis. *Current opinion in infectious diseases* **2009**, *Volume 22*, pp. 174.

45. Ryu, Y.J., Diagnosis of pulmonary tuberculosis: recent advances and diagnostic algorithms. *Tuberculosis respiratory diseases* **2015**, *Volume 78*, pp. 64-71. https://doi.org/10.4046/trd.2015.78.2.64.

46. Singhal, R.; Myneedu, V.P., Microscopy as a diagnostic tool in pulmonary tuberculosis. *International journal of mycobacteriology* **2015**, *Volume 4*, pp. 1-6. https://doi.org/10.1016/j.ijmyco.2014.12.006.

47. Engel, N.; Wachter, K.; Pai, M.; Gallarda, J.; Boehme, C.; Celentano, I.; Weintraub, R., Addressing the challenges of diagnostics demand and supply: insights from an online global health discussion platform. *BMJ Global Health* **2016**, *Volume 1*, pp. e000132. https://doi.org/10.1136/bmjgh-2016-000132.

48. Pantoja, A.; Fitzpatrick, C.; Vassall, A.; Weyer, K.; Floyd, K., Xpert MTB/RIF for diagnosis of tuberculosis and drug-resistant tuberculosis: a cost and affordability analysis. *European Respiratory Journal* **2013**, *Volume 42*, pp. 708-720. https://doi.org/10.1183/09031936.00147912.

49. Cazabon, D.; Pande, T.; Kik, S.; Van Gemert, W.; Sohn, H.; Denkinger, C.; Qin, Z.Z.; Waning, B.; Pai, M., Market penetration of Xpert MTB/RIF in high tuberculosis burden countries: a trend analysis from 2014-2016. *Gates open research* **2018**, *Volume 2*, pp. https://doi.org/10.12688/gatesopenres.12842.2.

50. Sakamoto, H.; Lee, S.; Ishizuka, A.; Hinoshita, E.; Hori, H.; Ishibashi, N.; Komada, K.; Norizuki, M.; Katsuma, Y.; Akashi, H., Challenges and opportunities for eliminating tuberculosis–leveraging political momentum of the UN high-level meeting on tuberculosis. *BMC public health* **2019**, *Volume 19*, pp. 76. https://doi.org/10.1186/s12889-019-6399-8.

51. Tanimura, T.; Jaramillo, E.; Weil, D.; Raviglione, M.; Lönnroth, K., Financial burden for tuberculosis patients in low-and middle-income countries: a systematic review. *European Respiratory Journal* **2014**, *Volume 43*, pp. 1763-1775. https://doi.org/10.1183/09031936.00193413.

52. Reed, J.L.; Basu, D.; Butzler, M.A.; McFall, S.M., XtracTB Assay, a Mycobacterium tuberculosis molecular screening test with sensitivity approaching culture. *Scientific reports* **2017**, *Volume 7*, pp. 3653. https://doi.org/10.1038/s41598-017-03930-3.

53. García-Basteiro, A.L.; DiNardo, A.; Saavedra, B.; Silva, D.; Palmero, D.; Gegia, M.; Migliori, G.B.; Duarte, R.; Mambuque, E.; Centis, R., Point of care diagnostics for tuberculosis. *Pulmonology* **2018**, *Volume 24*, pp. 73-85. https://doi.org/10.1016/j.rppnen.2017.12.002.

54. Bhusal, N.; Shrestha, S.; Pote, N.; Alocilja, E., Nanoparticle-Based Biosensing of Tuberculosis, an Affordable and Practical Alternative to Current Methods. *Biosensors* **2019**, *Volume 9*, pp. 1. https://doi.org/10.3390/bios9010001.

55. Gordillo-Marroquín, C.; Gómez-Velasco, A.; Sánchez-Pérez, H.; Pryg, K.; Shinners, J.; Murray, N.; Muñoz-Jiménez, S.; Bencomo-Alerm, A.; Gómez-Bustamante, A.; Jonapá-Gómez, L., Magnetic Nanoparticle-Based Biosensing Assay Quantitatively Enhances Acid-Fast Bacilli Count in Paucibacillary Pulmonary Tuberculosis. *Biosensors* **2018**, *Volume 8*, pp. 128. https://doi.org/10.3390/bios8040128.

56. Floyd, K.; Glaziou, P.; Houben, R.; Sumner, T.; White, R.; Raviglione, M., Global tuberculosis targets and milestones set for 2016–2035: definition and rationale. *The international journal of tuberculosis lung disease* **2018**, *Volume 22*, pp. 723-730. https://doi.org/10.5588/ijtld.17.0835.

57. Kristensen, A.K.B.; Thomsen, J.F.; Mikkelsen, S., A review of mercury exposure among artisanal small-scale gold miners in developing countries. *International archives of occupational environmental health* **2014**, *Volume 87*, pp. 579-590.

58. Hentschel, T.; Hruschka, F.; Priester, M. *Artisanal and Small-Scale Mining: Challenges and Opportunities*, 1st ed.; Inernational Institute for Environment and Development (IIED) and World Business Council for Sustainable Development: London, United Kingdom, 2003.

59. Cordy, P.; Veiga, M.M.; Salih, I.; Al-Saadi, S.; Console, S.; Garcia, O.; Mesa, L.A.; Velásquez-López, P.C.; Roeser, M., Mercury contamination from artisanal gold mining in Antioquia, Colombia: The world's highest per capita mercury pollution. *Science of the Total Environment* **2011**, *Volume 410*, pp. 154-160.

60. United Nations Environment Programme, Global mercury assessment 2013: Sources, emissions, releases and environmental transport. *UNEP Chemicals Branch, Geneva, Switzerland.* 2013.

61. Sengupta, P.; Banerjee, R.; Nath, S.; Das, S.; Banerjee, S., Metals and female reproductive toxicity. *Human experimental toxicology* **2015**, *Volume 34*, pp. 679-697.

62. Pizent, A.; Tariba, B.; Živković, T., Reproductive toxicity of metals in men. *Archives of industrial hygiene toxicology* **2012**, *Volume 63*, pp. 35-46.

63. Matta, G.; Gjyli, L., Mercury, lead and arsenic: impact on environment and human health. *Journal of chemical Pharmaceutical Sciences* **2016**, *Volume 9*, pp. 718-725.

64. Mahurpawar, M., Effects of heavy metals on human health. *Int. J. Res. Granthaalayah* **2015**, *Volume 530*, pp. 1-7.

65. Bridges, C.C.; Zalups, R.K., The aging kidney and the nephrotoxic effects of mercury. *Journal of Toxicology Environmental Health, Part B* **2017**, *Volume 20*, pp. 55-80.

66. Ghaedi, M.; Reza Fathi, M.; Shokrollahi, A.; Shajarat, F., Highly selective and sensitive preconcentration of mercury ion and determination by cold vapor atomic absorption spectroscopy. *Analytical Letters* **2006**, *Volume 39*, pp. 1171-1185.

67. Yavuz, E.; Tokalıoğlu, Ş.; Patat, Ş., Magnetic dispersive solid phase extraction with graphene/ZnFe2O4 nanocomposite adsorbent for the sensitive determination of mercury in water and fish samples by cold vapor atomic absorption spectrometry. *Microchemical Journal* **2018**, *Volume 142*, pp. 85-93.

68. Jones, R.; Jacobson, M.; Jaffe, R.; West-Thomas, J.; Arfstrom, C.; Alli, A., Method development and sample processing of water, soil, and tissue for the analysis of total and organic mercury by cold vapor atomic Fluorescence spectrometry. *Oceanographic Literature Review* **1996**, *Volume 2*, pp. 192.

69. Yu, L.-P.; Yan, X.-P., Flow injection on-line sorption preconcentration coupled with cold vapor atomic fluorescence spectrometry and on-line oxidative elution for the determination of trace mercury in water samples. *Atomic spectroscopy* **2004**, *Volume 25*, pp. 145-153.

70. Rodríguez-Reino, M.P.; Rodríguez-Fernández, R.; Peña-Vázquez, E.; Domínguez-González, R.; Bermejo-Barrera, P.; Moreda-Piñeiro, A., Mercury speciation in seawater by liquid chromatography-inductively coupled plasma-mass spectrometry following solid phase extraction pre-concentration by using an ionic imprinted polymer based on methyl-mercury–phenobarbital interaction. *Journal of Chromatography A* **2015**, *Volume 1391*, pp. 9-17.

71. Ma, S.; He, M.; Chen, B.; Deng, W.; Zheng, Q.; Hu, B., Magnetic solid phase extraction coupled with inductively coupled plasma mass spectrometry for the speciation of mercury in environmental water and human hair samples. *Talanta* **2016**, *Volume 146*, pp. 93-99.

72.   Barton, J.; García, M.B.G.; Santos, D.H.; Fanjul-Bolado, P.; Ribotti, A.; McCaul, M.; Diamond, D.; Magni, P., Screen-printed electrodes for environmental monitoring of heavy metal ions: a review. *Microchimica Acta* **2016**, *Volume 183*, pp. 503-517.

73.   Kabir, K.M.; Sabri, Y.M.; Lay, B.; Ippolito, S.J.; Bhargava, S.K., A silver electrode based surface acoustic wave (SAW) mercury vapor sensor: a physio-chemical and analytical investigation. *RSC Advances* **2016**, *Volume 6*, pp. 36362-36372.

74.   Lai, C.; Qin, L.; Zeng, G.; Liu, Y.; Huang, D.; Zhang, C.; Xu, P.; Cheng, M.; Qin, X.; Wang, M., Sensitive and selective detection of mercury ions based on papain and 2, 6-pyridinedicarboxylic acid functionalized gold nanoparticles. *RSC Advances* **2016**, *Volume 6*, pp. 3259-3266.

75.   Noël, M.; Christensen, J.R.; Spence, J.; Robbins, C.T., Using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) to characterize copper, zinc and mercury along grizzly bear hair providing estimate of diet. *Science of the Total Environment* **2015**, *Volume 529*, pp. 1-9.

76.   Habte, G.; Hwang, I.M.; Kim, J.S.; Hong, J.H.; Hong, Y.S.; Choi, J.Y.; Nho, E.Y.; Jamila, N.; Khan, N.; Kim, K.S., Elemental profiling and geographical differentiation of Ethiopian coffee samples through inductively coupled plasma-optical emission spectroscopy (ICP-OES), ICP-mass spectrometry (ICP-MS) and direct mercury analyzer (DMA). *Food chemistry* **2016**, *Volume 212*, pp. 512-520.

77.   Cui, L.; Wu, J.; Ju, H., Electrochemical sensing of heavy metal ions with inorganic, organic and bio-materials. *Biosensors Bioelectronics* **2015**, *Volume 63*, pp. 276-286.

78.   Bansod, B.; Kumar, T.; Thakur, R.; Rana, S.; Singh, I., A review on various electrochemical techniques for heavy metal ions detection with different sensing platforms. *Biosensors Bioelectronics* **2017**, *Volume 94*, pp. 443-455.

79.   Gao, C.; Huang, X.-J., Voltammetric determination of mercury (II). *TrAC Trends in Analytical Chemistry* **2013**, *Volume 51*, pp. 1-12.

80.   Mauter, M.S.; Elimelech, M., Environmental applications of carbon-based nanomaterials. *Environmental science technology* **2008**, *Volume 42*, pp. 5843-5859.

81.   Crevillen, A.G.; Escarpa, A.; García, C.D. *Carbon-based Nanomaterials in Analytical Chemistry*, 1st ed.; Royal Society of Chemistry: United Kingdom, 2018.

82.   Xie, F.; Yang, M.; Jiang, M.; Huang, X.-J.; Liu, W.-Q.; Xie, P.-H., Carbon based nanomaterials-A promising electrochemical sensor toward persistent toxic substance. *Trends in Analytical Chemistry* **2019**, *Volume* pp. 115624.

83.   Gong, J.; Zhou, T.; Song, D.; Zhang, L.; Hu, X., Stripping voltammetric detection of mercury (II) based on a bimetallic Au– Pt inorganic– organic hybrid nanocomposite modified glassy carbon electrode. *Analytical chemistry* **2009**, *Volume 82*, pp. 567-573.

84.   Ghanei-Motlagh, M.; Taher, M.A.; Heydari, A.; Ghanei-Motlagh, R.; Gupta, V.K., A novel voltammetric sensor for sensitive detection of mercury (II) ions using glassy carbon electrode modified with graphene-based ion imprinted polymer. *Materials Science Engineering: C* **2016**, *Volume 63*, pp. 367-375.

85.   Pokhrel, L.R.; Ettore, N.; Jacobs, Z.L.; Zarr, A.; Weir, M.H.; Scheuerman, P.R.; Kanel, S.R.; Dubey, B., Novel carbon nanotube (CNT)-based ultrasensitive sensors for trace mercury (II) detection in water: a review. *Science of the Total Environment* **2017**, *Volume 574*, pp. 1379-1388.

86.   Xing, H.; Xu, J.; Zhu, X.; Duan, X.; Lu, L.; Wang, W.; Zhang, Y.; Yang, T., Highly sensitive simultaneous determination of cadmium (II), lead (II), copper (II), and mercury (II) ions on N-doped graphene modified electrode. *Journal of Electroanalytical Chemistry* **2016**, *Volume 760*, pp. 52-58.

87.   Choi, S.-M.; Kim, D.-M.; Jung, O.-S.; Shim, Y.-B., A disposable chronocoulometric sensor for heavy metal ions using a diaminoterthiophene-modified electrode doped with graphene oxide. *Analytica chimica acta* **2015**, *Volume 892*, pp. 77-84.

88.   Huang, X.-J.; Chen, X.; Yang, M. *Persistent Toxic Substance Monitoring: Nanoelectrochemical Methods*, 1st ed.; John Wiley & Sons: Weinheim, Germany, 2018.

89.   Duarte, K.; Justino, C.I.; Freitas, A.C.; Gomes, A.M.; Duarte, A.C.; Rocha-Santos, T.A., Disposable sensors for environmental monitoring of lead, cadmium and mercury. *TrAC Trends in Analytical Chemistry* **2015**, *Volume 64*, pp. 183-190.

90. Lin, J.; Peng, Z.; Liu, Y.; Ruiz-Zepeda, F.; Ye, R.; Samuel, E.L.; Yacaman, M.J.; Yakobson, B.I.; Tour, J.M., Laser-induced porous graphene films from commercial polymers. *Nature communications* **2014**, *Volume 5*, pp. 5714.

91. Tehrani, F.; Bavarian, B., Facile and scalable disposable sensor based on laser engraved graphene for electrochemical detection of glucose. *Scientific reports* **2016**, *Volume 6*, pp. 27975.

92. Nayak, P.; Kurra, N.; Xia, C.; Alshareef, H.N., Highly efficient laser scribed graphene electrodes for on‐chip electrochemical sensing applications. *Advanced Electronic Materials* **2016**, *Volume 2*, pp. 160085. 10.1002/aelm.201600185.

93. Vanegas, D.; Patiño, L.; Mendez, C.; Oliveira, D.; Torres, A.; Gomes, C.; McLamore, E., Laser Scribed Graphene Biosensor for Detection of Biogenic Amines in Food Samples Using Locally Sourced Materials. *Biosensors* **2018**, *Volume 8*, pp. 42.

94. Garland, N.T.; McLamore, E.S.; Cavallaro, N.D.; Mendivelso-Perez, D.; Smith, E.A.; Jing, D.; Claussen, J.C., Flexible Laser-Induced Graphene for Nitrogen Sensing in Soil. *ACS applied materials interfaces* **2018**, *Volume 10*, pp. 39124-39133.

95. Yang, K.; Wang, J.; Chen, X.; Zhao, Q.; Ghaffar, A.; Chen, B., Application of graphene-based materials in water purification: from the nanoscale to specific devices. *Environmental Science: Nano* **2018**, *Volume 5*, pp. 1264-1297.

96. Taguchi, M.; Schwalb, N.; Rong, Y.; Vanegas, D.; Garland, N.; Tan, M.; Yamaguchi, H.; Claussen, J.; McLamore, E., Pulsed: Pulsed sonoelectrodeposition of fractal nanoplatinum for enhancing amperometric biosensor performance. *Analyst* **2016**, *Volume 141*, pp. 3367-3378.

97. Abdelbasir, S.; El-Sheikh, S.; Morgan, V.; Schmidt, H.; Casso-Hartmann, L.; Vanegas, D.; Velez-Torres, I.; McLamore, E., Graphene-anchored cuprous oxide nanoparticles from waste electric cables for electrochemical sensing. *Sustainable Chemistry Engineering* **2018**, *Volume 6*, pp. 12176-12186.

98. Jenkins, D.M.; Lee, B.E.; Jun, S.; Reyes-De-Corcuera, J.; McLamore, E.S., ABE-Stat, a Fully Open-Source and Versatile Wireless Potentiostat Project Including Electrochemical Impedance Spectroscopy. *Journal of The Electrochemical Society* **2019**, *Volume 166*, pp. B3056-B3065.

99. World Health Organization, Mercury in Drinking-water: Background document for development of WHO Guidelines for Drinking-water Quality. 2005.

100. World Health Organization, Exposure to mercury: a major public health concern. *WHO, Public Health Environment international.* 2007.

101. Taueg, C.; Sanfilippo, D.; Rowens, B.; Szejda, J.; Hesse, J., Acute and chronic poisoning from residential exposures to elemental mercury-michigan, 1989–1990. *Journal of Toxicology: Clinical Toxicology* **1992**, *Volume 30*, pp. 63-67.

102. Bernhoft, R.A., Mercury toxicity and treatment: a review of the literature. *Journal of environmental public health* **2012**, *Volume 2012*.

103. Castilhos, Z.C.; Rodrigues-Filho, S.; Rodrigues, A.P.C.; Villas-Bôas, R.C.; Siegel, S.; Veiga, M.M.; Beinhoff, C., Mercury contamination in fish from gold mining areas in Indonesia and human health risk assessment. *Science of the Total Environment* **2006**, *Volume 368*, pp. 320-325.

104. Castilhos, Z.; Rodrigues-Filho, S.; Cesar, R.; Rodrigues, A.P.; Villas-Bôas, R.; de Jesus, I.; Lima, M.; Faial, K.; Miranda, A.; Brabo, E., Human exposure and risk assessment associated with mercury contamination in artisanal gold mining areas in the Brazilian Amazon. *Environmental Science Pollution Research* **2015**, *Volume 22*, pp. 11255-11264.

105. Nakazawa, K.; Nagafuchi, O.; Kawakami, T.; Inoue, T.; Yokota, K.; Serikawa, Y.; Cyio, B.; Elvince, R., Human health risk assessment of mercury vapor around artisanal small-scale gold mining area, Palu city, Central Sulawesi, Indonesia. *Ecotoxicology environmental safety* **2016**, *Volume 124*, pp. 155-162.

106. Petroczi, A.; Naughton, D., Mercury, cadmium and lead contamination in seafood: A comparative study to evaluate the usefulness of Target Hazard Quotients. *Food Chemical Toxicology* **2009**, *Volume 47*, pp. 298-302.

107. Telemedicine Predicted in 1925. Available online: https://www.smithsonianmag.com/history/telemedicine-predicted-in-1925-124140942/ (accessed on November 20 2019).

108. Arute, F.; Arya, K.; Babbush, R.; Bacon, D.; Bardin, J.C.; Barends, R.; Biswas, R.; Boixo, S.; Brandao, F.G.; Buell, D.A., Quantum supremacy using a programmable superconducting processor. *Nature* **2019**, *Volume 574*, pp. 505-510.

109. Coase, R., The Nature of the Firm. Economica, New Series, 4 (16), 386-405. *Article first published online* **1937**, *Volume 19*, pp. 10.1111/j.1468-0335.1937.tb00002.x.

110. Kay, N.M., Coase and the Contribution of 'The Nature of the Firm'. *Managerial Decision Economics* **2015**, *Volume 36*, pp. 44-54.

111. Kim, J.Y., Rich and poor: opportunities and challenges in an age of disruption. *World Bank* 2018. 10.1596/31119.

112. Tonda, E.; Susan, C., Technology Challenges and tools for the implementation of the water-related sustainable development goals and targets. *2015 UN-Water Annual International Zaragoza Conference* 2015.

113. Banerjee, A.V. and Duflo, E. *Poor economics: A radical rethinking of the way to fight global poverty*. 1st ed.; Public Affairs: New York, United States of America, 2011. ISBN: 978-1-58648-798-0

114. Mikami, M., Methodological divergence between coase and williamson in the history of transaction cost economics. *Economic Journal of Hokkaido University* **2011**, *Volume 40*, pp. 41-57.

115. Klein, S.; Frazier, G.L.; Roth, V.J., A transaction cost analysis model of channel integration in international markets. *Journal of Marketing research* **1990**, *Volume 27*, pp. 196-208.

116. Pentland, A.; Fletcher, R.; Hasson, A., Daknet: Rethinking connectivity in developing nations. *Computer* **2004**, *Volume 37*, pp. 78-83. 10.1109/MC.2004.1260729.

117. Martinez, A.; Villarroel, V.; Seoane, J.; del Pozo, F., Analysis of information and communication needs in rural primary health care in developing countries. *IEEE transactions on Information Technology in Biomedicine* **2005**, *Volume 9*, pp. 66-72. 10.1109/TITB.2004.842411.

118. Pimenidis, E.; Sideridis, A.B.; Antonopoulou, E., Mobile devices and services: bridging the digital divide in rural areas. *International Journal of Electronic Security Digital Forensics* **2009**, *Volume 2*, pp. 424-434.

119. James, J., Mechanisms of access to the Internet in rural areas of developing countries. *Telematics Informatics* **2010**, *Volume 27*, pp. 370-376. https://doi.org/10.1016/j.tele.2010.02.002.

120. Llanto, G.M.; Fukui, R., Innovations in Microfinance in Southeast Asia, PIDS Discussion Paper Series, No. 2003-11, Philippine Institute for Development Studies (PIDS), Makati City. 2003. *PIDS Discussion Paper Series*

121. Nenova, T.; Niang, C.T. *Bringing Finance to Pakistan's poor: Access to Finance for Small enterprises and the Underserved*, 1st ed.; The World Bank: Washington DC, United States of America, 2009.

122. Bernardi, L.; Sarma, S.; Traub, K. *The inversion factor: How to thrive in the IoT economy*, 1st ed.; MIT Press: London, England, 2017.

123. Davis, F.D., A technology acceptance model for empirically testing new end-user information systems: Theory and results. Ph.D. thesis. Ph.D. in Management-Massachusetts Institute of Technology, Massachusetts, United States of America, December 20 1985.

124. Rogers, E.M., A prospective and retrospective look at the diffusion model. *Journal of health communication* **2004**, *Volume 9*, pp. 13-19.

125. Williams, M.D.; Rana, N.P.; Dwivedi, Y.K., The unified theory of acceptance and use of technology (UTAUT): a literature review. *Journal of Enterprise Information Management* **2015**, *Volume 28*, pp. 443-488. 10.1108/JEIM-09-2014-0088.

126. Johnson, M.P.; Zheng, K.; Padman, R., Modeling the longitudinality of user acceptance of technology with an evidence-adaptive clinical decision support system. *Decision Support Systems* **2014**, *Volume 57*, pp. 444-453. 10.1016/j.dss.2012.10.049.

127. Campbell, J.I.; Aturinda, I.; Mwesigwa, E.; Burns, B.; Haberer, J.E.; Bangsberg, D.R.; Holden, R.J.; Ware, N.C.; Siedner, M.J., The Technology Acceptance Model for Resource-Limited Settings (TAM-RLS): a novel framework for mobile health interventions targeted to low-literacy end-users in resource-limited settings. *AIDS Behavior* **2017**, *Volume 21*, pp. 3129-3140. 10.1007/s10461-017-1765-y.

128. Sunny, S.; Patrick, L.; Rob, L., Impact of cultural values on technology acceptance and technology readiness. *International Journal of Hospitality Management* **2019**, *Volume 77*, pp. 89-96. https://doi.org/10.1016/j.ijhm.2018.06.017.

129. Mathieson, K., Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. *Information systems research* **1991**, *Volume 2*, pp. 173-191.

130. Wallace, E.A.; Schumann, J.H.; Weinberger, S.E., Ethics of Commercial Screening Tests. *Annals of Internal Medicine* **2013**, *Volume 158*, pp. 500-500. doi:10.231/JIM.0b013e318210eeb0.

131. Hersh, M., Technology change, technology transfer and ethics. University of Glasgow, Vienna, Austria. 2001.

987   132.   Street, J. *Politics and technology*, 1st ed.; The Guilford Press: New York, United Stated, 1992.
988   133.   Guidelines for health care equipment donations. Available online:
989   https://www.who.int/medical_devices/publications/en/Donation_Guidelines.pdf   (accessed
990   on November 05 2019).
991   134.   Perry, L.; Malkin, R., Effectiveness of medical equipment donations to improve health systems:
992   how much medical equipment is broken in the developing world? *Medical & Biological*
993   *Engineering & Computing* **2011**, *Volume 49*, pp. 719-722. https://doi.org/10.1007/s11517-011-0786-
994   3.
995

996

997

Can "SNAPS" and/or "PAPPU" make a difference for the masses?

The New York Times

Economic Incentives Don't Always Do What We Want Them To

By Esther Duflo and Abhijit Banerjee
The authors were just awarded the Nobel Prize in economics.

Oct. 26, 2019

f   ▼   ✉   ➔   🔖   359

P3

Porous Pareto Partition:

[1]Democratization of [2]Distributed [3]Data may or may not catalyze the global [4]Dissemination of [5]Digital [6]Dividends

D[6] is an unfinished essay (op-ed) by Shoumen Palit Austin Datta *et al* ◆ Massachusetts Institute of Technology and Massachusetts General Hospital, Harvard Medical School (*shoumen@mit.edu* & *sdatta8@mgh.harvard.edu*)
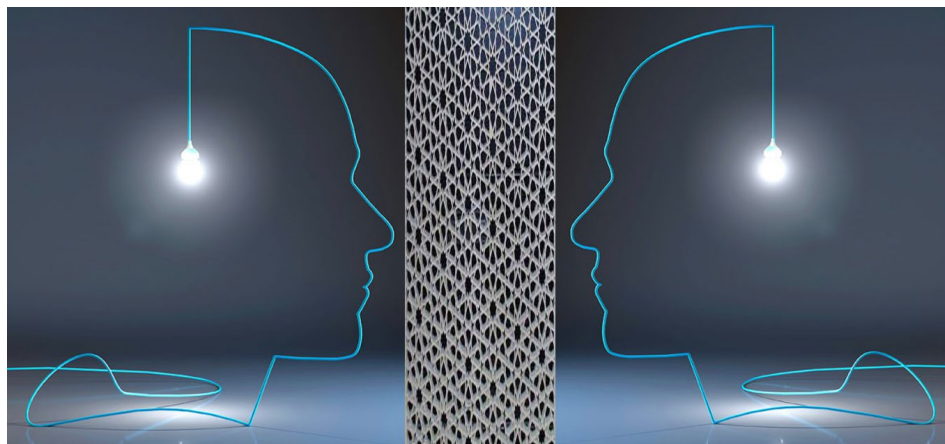
ABSTRACT

P3 is a petri dish brimming with questions. This essay does not provide answers but suggestions to explore. This essay may not instruct, educate or teach, but may swing the proverbial pendulum with sufficient motion to discuss business to consumer services. The reader may ponder about the amorphous questions or wonder in confusion, ambiguity and uncertainty. This essay breaks with the *status quo* and indulges orthogonal, non-linear and asymmetric information arbitrage. This essay is not an attempt to be correct or right. This essay is a seed, sterile unless cultivated or you can bury it (but it may grow). This essay aspires to inform that tools and data related to the affluent world are **_not_** a template to be "copied" or applied to systems in the remaining (80%) parts of the world with economic constraints and cultural preferences. We need different thinking that resists the inclination of the affluent 20% of the world to treat the rest of the world (80% of the population) just as a market. The 80/20 distribution evokes the Pareto[1] theme in this essay. On reflection (see "END NOTE"), it is unclear what this essay contributes.

CONTEXT

Since 1999, the concept of the internet of things (IoT) was nurtured as a marketing term[2] which may have succinctly captured the idea of data about objects stored on the internet[3] in the networked physical world. The idea evolved while transforming the use of RFID (radio frequency identification) where an alphanumeric unique identifier (64-bit EPC[4] or electronic product code) was stored on the chip (tag[5]) but the volume of raw data was stored on the internet, yet inextricably and uniquely linked via the EPC, in a manner resembling the structure of internet protocols[6] (64-bit IPv4 and 128-bit IPv6[7]). IoT and later, *cloud of data[8]*, were metaphors for ubiquitous connectivity and concepts originating from ubiquitous computing, a term introduced by Mark Weiser[9] in 1998. The underlying importance of data from connected objects and processes usurped the term big data[10] and then twisted the sound bites to create the artificial myth of "Big Data" sponsored and accelerated by consulting companies. The global drive to get ahead of the "Big Data" tsunami, flooded both businesses and governments, big and small. The chatter about big data garnished with dollops of AI became parlor talk among fish mongers[11] and gold miners, inviting the sardonicism of doublespeak, peppered in this essay.

Much to the chagrin of the thinkers, the laissez-faire approach to IoT percolated by the tinkerers over-shadowed hard facts. The "quick & dirty" anti-intellectual chaos adumbrated the artefact-fueled exploding frenzy for new revenue from "IoT Practice" which spawned greed in the consulting[12] world. The cacophony of IoT in the market[13] is a result of that unstoppable transmutation of disingenuous tabloid fodder to veritable truth, catalyzed by pseudo-science hacks, social gurus and glib publicity campaigns to drum up draconian "dollar-sign-dangling" predictions[14] about "trillions of things connected to the internet" to feed mass hysteria, to bolster consumption. Few ventured to correct the facts and point out that *connectivity without discovery* is a diabolical tragedy of egregious errors. Even fewer recognized that the idea of IoT is *not a point* but an **ecosystem**, where collaboration adds value.

The corporate orchestration of the digital by design metaphor of IoT was warped solely to create demand for sales by falsely amplifying the lure of increasing performance, productivity and profit, far beyond the potential digital transformation could deliver by embracing the rational principles of IoT.

Ubiquitous connectivity is associated with high cost of products (capex or capital expense) but extraction of "value" to generate ROI (return on investment) rests on the ability to implement SARA, a derivative of the PEAS paradigm (see Fig 6 and 7). SARA - Sense, Analyze, Respond, Actuate – is not a linear concept. Data and decisions necessary for SARA makes the conceptual illustration more akin to The Sara Cycle, perhaps best illustrated by the analogy to the Krebs[15] Cycle, an instance of bio-mimicry. Data and decisions constantly influence, optimize, re-configure, and change the parameters associated with, *when* to sense, *what* to analyze, *how* to respond and *where* to actuate or auto-actuate. Combining SARA with the metaphor of IoT by design may help to ask these questions, with precision and accuracy.

It is hardly necessary to over-emphasize the value of the correct questions for each element of SARA in a matrix of connected objects, relevant entities which can be discovered, distributed nodes, related processes and desired outcomes. Strategic inclusion of SARA guides key performance indicators. Lucidity and clarity of thoughtful integration of digital by design idea is key to reconfiguring operations management. Execution and embedding SARA is not a systems integration task but rather a fine-tuned *synergistic* integration based on the *weighted combination of dependencies* in the SARA matrix. Failure to grasp the role of data and semantics of queries, in the context of KPI (key performance indicators) may increase transaction costs, reduce the value proposition for customers and obliterate ROI or profitability.

Figure 0 - From the annals[16] of the march of unreason: *Internet of things: $8.9 trillion market in 2020, 212 billion connected things.* It is blasphemous and heretical to suggest that this a *research*[17] outcome.

This essay meanders, not always aimlessly, around discussions involving data and decision. It also oscillates, albeit asynchronously, between a broad spectrum of haphazard realities or "dots" which may be more about esoteric analysis rather than focusing on delivering real-world value. In part, this discussion questions the barriers to the rate of diffusion of technologies in underserved communities. Can implementing *simple* tools act as affordable catalysts? Can it lift the quality of life, in less affluent societies, by enabling meaningful use of data, perhaps small data, at the right time, at the lowest cost?

The extremely non-linear business of delivering tools and technologies makes it imperative to consider the trinity of systems integration, standards and interoperability. We advocate that businesses may wish to gradually dis-engage with the product mind-set (sensors, hardware, software) and engage in the *ecosystem* necessary to deliver **_services_** to communities. The delivery of service to the end-user must be synergized. Hence, system integration may be a subset of synergistic integration. But, before we can view this "whole" it is better to understand the coalition of cyber (data) with the physical (parts). In many ways, this discussion is about cyberphysical systems (CPS) but not for lofty purposes, such as landing on Mars, but for simple living, on Earth.

MODELS IN THE BACKGROUND

Because it may be difficult to grasp the whole, we tend to focus on the part, and parts, closest to our comfort zone, in our area of interest. This reductionist approach may be necessary *ab initio* but rarely yields a solution, *per se*. Reconstruction requires synthesis and synergy, the global glue which underlies mass adoption and diffusion, of tools, in an age of integration, which, itself, is a khichuri[18] of parts, some known (industrial age, information age, systems age) and others, parts unknown.

Divide and conquer still remains a robust adage. It may be the philosophical foundation of reductionism. The latter has rewarded us with immense gains in knowledge and the wisdom as to why this *modus operandi* is *sine qua non*. For example, the pea plant (*Pisum sativum*) unleashed the cryptic principles of genetics[19] and unicellular bacteria shed light on normal physiological underpinnings of feedback control[20] common in genetic circuits as well as regulatory networks for maintenance and optimization of biological homeostasis, quintessential for health and healthcare in humans and animals. Cancer biology was transformed by Renato Dulbecco[21] by **_reducing_** the multi-factorial complexity of human cancer research to focus on a **_single gene model_** (SV40 large T-antigen) from Papova viruses.

Biomimicry also inspired the creation of better machines and systems[22], using the principles and practice of control theory borrowed from science, strengthened by mathematics and successfully integrated with design and manufacturing, by engineers. An early convergence[23] of control theory with communication may be found in the 1948 treatise "Cybernetics" by Norbert Wiener[24] (who may have borrowed[25] the word "cybernétique" proposed by the French physicist and mathematician André-Marie Ampère[26] to design the then non-existent science of process control).

In other examples of 'divide and conquer' the theoretical duo "Alice and Bob" is at the core[27] of cryptography[28] as well as the game theoretic[29] approach[30] to "prisoner's dilemma" which has influenced business strategies[31] and now it is spilling over to knowledge graph[32] databases. The simple concept of a lone travelling salesman proposed by Euler in 1759 appears to have evolved[33] as the bread and butter of most optimization engines, which, when considered together with data and information, continues to improve decision support systems (DSS) in manufacturing, retail, transportation, logistics[34] and omnipresent supply chain[35] networks, almost in every vertical.
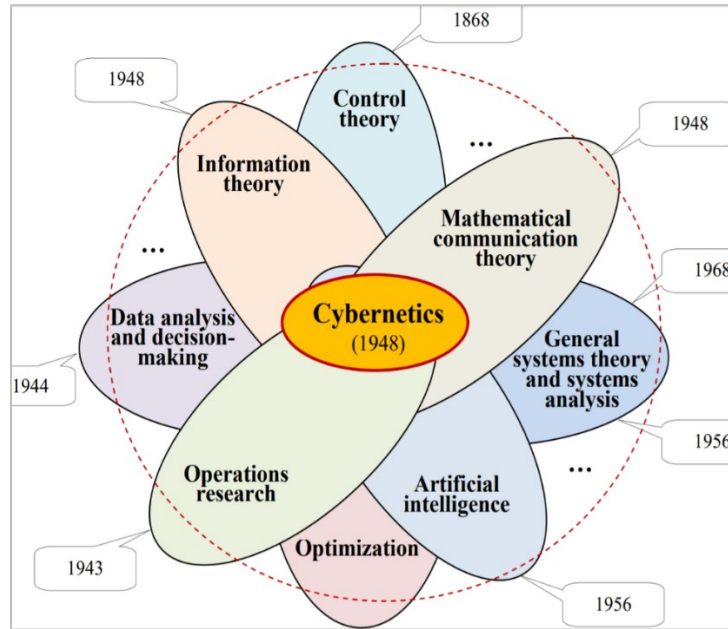
Figure 1 - A Century of Convergence ◆ The Composition and Structure of Cybernetics [22]

The purpose of these disparate examples are to emphasize the notion that there are fundamental units of activity or models or set(s) of patterns or certain basic behavioral criteria (for lack of a better descriptive term) that underlie most actions and reactions. When taken apart or sufficiently reduced, we may observe these as isolated units or patterns or models of rudimentary entities. When combined, these simple models/units/patterns/elements can generate an almost unlimited variety of system behaviors observed on grand scales. When viewing the massive scale of systems from the "top" it may be quite counter-intuitive to imagine that the observed manifestations are due to a few or a relatively small group of universal 'truths' which we refer to as models, units, rules, logic, patterns, elements or behaviors. To further illustrate this perspective, consider petals (flowers), pineapple (fruit) and pyramids. The variation between and within these three very different examples may boil down to Fibonacci[36] numbers, fractal[37] dimensions and the Golden[38] Ratio[39] in some form, or the other. In another vein, the number, eight, seems to be central to atoms (octet) and an integral part of the Standard Model in physics (octonions[40]). Number 8 is revered by the Chinese due to its link with words synonymous with wealth and fortune (fa).

If one is still unconvinced and remain skeptical that small sets of underlying elements, generally, may be responsible, albeit in part, for the "big things" we consider diverse, then the "killer" example is that of nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), made up of only five sub-units or molecules (Adenine, Guanine, Cytosine, Thymine, Uracil). DNA and RNA serves as the blueprint for all humans, animals, plants, bacteria and viruses, that may ever exist. The infinite diversity of multi-cellular[41] and uni-cellular organisms, whose creation is instructed by a *combination of these 5 molecules* in DNA and RNA, may vastly exceed 5×10^{30} (5,000,000,000,000,000,000,000,000,000,000[42]). The known exception to the DNA-RNA dogma may be the case for prions[43] which uses proteins[44] as the transmissible macromolecule.

Parallel examples can be drawn from physical sciences. Large scale system behaviors can be reduced and mapped to simple models. Combination of these simple models, with widely different microscopic details, applies to, and generates, large set of possible systems[45] and system of systems. Another example of "hidden complementarities" emerged from cryptic mathematical bridge embedded in natural sciences. It is now established that eigenvectors may be computed[46] using information about eigenvalues. Students are still taught that eigenvectors and eigenvalues are independent, and must be calculated separately starting from rows and columns of the matrix. Mathematicians authored papers in related fields[47] yet none "connected the dots" between eigenvectors and eigenvalues. The insight that eigenvalues of the minor matrix encode hidden information may not be entirely new[48] but was neither understood nor articulated. The relationship of centuries-old mathematical objects[49] ultimately came from physicists. Nature inspires mathematical thinking because mathematics thrives when connected to nature. Grasping these connections enables humans to create tools to mimic nature (bio-mimicry).
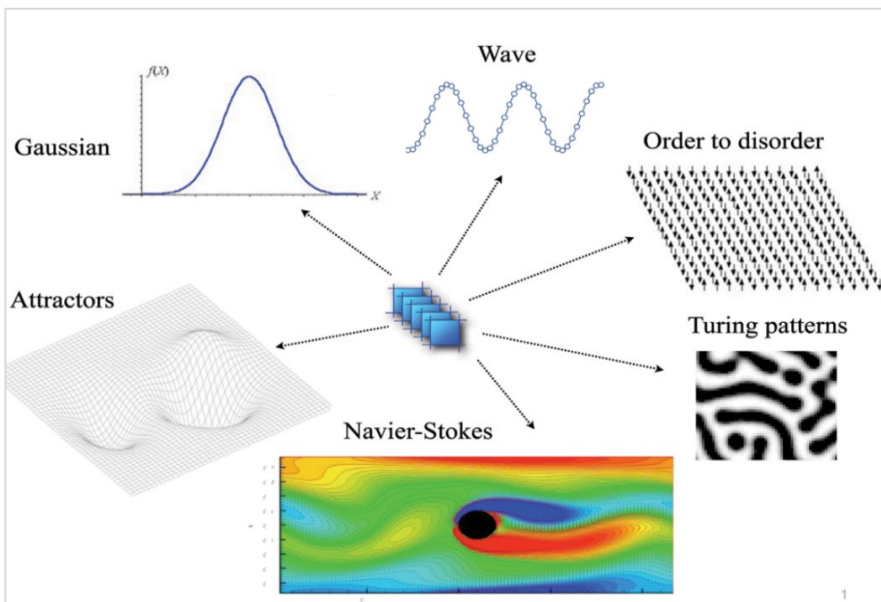
Figure 2 – Only a few models may capture the behavior of a wide range of systems, underlies the idea of universality[26] (models illustrated in this figure: Gaussian distribution, wave motion, order to disorder transitions, Turing patterns, fluid flow described by Navier-Stokes equations, and attractor dynamics).

PROBLEM SPACE: ARE WE ASKING THE CORRECT QUESTIONS?

The lengthy and winding preface [to the problem at hand (D_6)] is presented to substantiate the opinion that there may be a disconnect between the volume of data we have generated as a result of the "information age" versus the lack lustre gains in performance, as estimated by the productivity[50] index. Currently, we may have ~2.7 zettabytes[51] (2.7 billion terabytes) of data, but some estimates claim as much as 33 zettabytes[52] of data at hand (2018), and predicts it to reach 175 zettabytes circa 2025.

The deluge of data as a result of "information technology" is far greater in magnitude than the diffusion of electricity[53] a century ago. Productivity increases due to the introduction of electricity and IT offer economic parallels[54] but based on the magnitude of change, the short fall (in productivity) cannot be brushed aside by attributing the blame to mismeasurement explanations[55] for the sluggish[56] pace. Extrapolating measurements using the tools of classical productivity[57] to determine the impact of IT and influence of data is certainly fraught with problems[58] yet the incongruencies alone cannot explain the shrinkage. In socio-economic terms, there is a growing chasm between IT and data/information versus productivity, improvement in quality of life, labor, compensation[59] and increases in living standards.
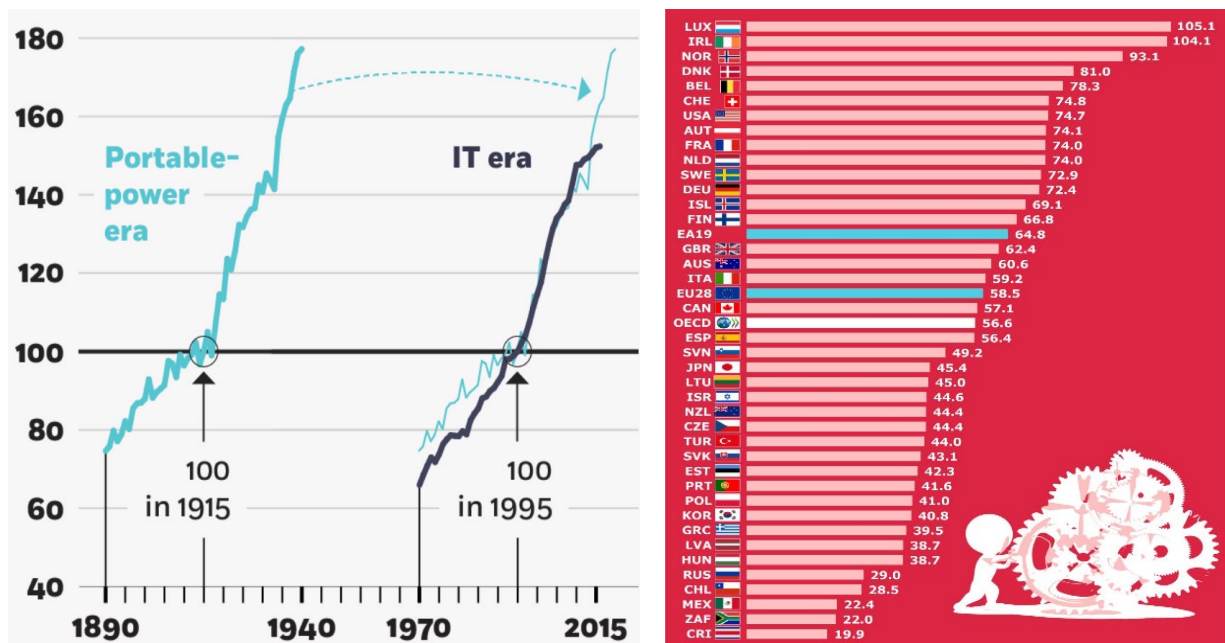
Figure 3 – (LEFT) Labor-Productivity Index [60]: Has data failed to deliver? IT was billed as the bridge between *the haves* and *the have-nots.* General process technologies takes ~25 years to reach market adoption [51]. (RIGHT) Labor Productivity[61] (OECD, 2018) is yet another example how the arithmetic of productivity (ratio between volume of output vs input) is misguided, misdiagnosed, mismeasured and misused as a metric of economic realities. Making Mexico (22.4) appear to be one fifth as "productive" as Ireland (104.1) suggests formulaic manipulations[62] (GDP per hour worked, current prices, PPP).

Despite trillions of dollars invested in data, digital transformation and other IT tools[63] (big data, AI, blockchain) the perforated return on investment[64] increasingly points to massive[65] waste. One reason for this "waste" may be due to use of models of data where errors are aggregated under a generalized[66] form or variations[67] of the normal (homoskedastic) distribution.

Heteroskedasticity was addressed[68] using ARCH[69] (autoregressive conditional heteroskedasticity[70]) and GARCH[71] models[72] (generalized ARCH). The use[73] of these proven techniques[74] for time series data (for example, sensor data showing water temperature in marine aquaponics[75] or cold chain[76] temperature log of vaccine package during transportation) in financial[77] econometrics[78] may be extended. Applications in predictive[79] modeling and forecasting[80] techniques may wish to adopt these econometric tools (GARCH) as a standard, whenever time series data is used (for example, supply chain[81] management, sensor data in health), but only *if* there is sufficient data (volume) to meet the statistical rigor necessary for successful error correction.

Perhaps it is best to limit the post-mortem analysis of IT failures, snake-oil sales of AI[82] and other debacles. Let us observe from this discussion that in the domain of data, and extraction of value from data to inform decisions, the tools of transforming data to inform decision may benefit from **re**-viewing the processes and technologies with "new" eyes. Above all, we must ask, often, if we are pursuing the correct questions, if the tools are appropriate and rigorous. The productivity gap and reports of corporate waste are "sign-posts" on the road ahead, except that the signage is in the incorrect direction, with respect to the intended destination, that is, profit and performance.

SOLUTIONS APPROACH – THE ELUSIVE QUEST TO BUILD BRIDGES BETWEEN DATA AND DECISIONS

There are no novel proposed solutions in this essay, only new commentary about *approaches* to solutions. The violent discord between volume of data versus veracity of decisions appears to be one prominent reason why the productivity gap may widen to form a chasm. The "background" section discussed how the reductionist approach points to simple models or underlying units or key elements, which, when combined, in some form, by some rules or logic, may generate large scale systems.

Data models[83] for DBMS are very different from models in data. Pattern mining[84] from data[85] is a time-tested tool. What new features can we uncover or learn about data, from patterns? What simpler models or elements are cryptic in data? Are these the correct questions? *If* there are simpler models or patterns in some types of data, can we justify extrapolating these models and *patterns* as a *general feature* of the data?

We have been mining for patterns and models (clustering, classification, categorization, principal component analysis) for decades, why haven't we found simpler models or patterns, yet? Are we using the wrong tools or wrong approaches or looking at the wrong places? How rational are we in our search for these general/simple models in view of the fact that models of data from retail or manufacturing or health clinics *should* be quite different?

The lowest common denominator of simple/general models/patterns may not be an ingredient for building that experimental "thought" bridge. Increasing volume of data could help GARCH tools but appears to be on a slippery slope in terms of quality with respect to *informing* decision support systems and/or the veracity of decisions (output). Baseline data models/patterns as denominators from grocery shopping or dry wall manufacturing or mental health clinics **should** be different. In lieu of "universal" common denominators, we may create repertoires of domain-specific common denominators. Then, a comparative analysis between common denominators of retail grocery shopping model from Boston versus Beijing may reveal the spectrum of nutritional behaviors. If linked to eating habits, perhaps we can extrapolate its *influence* on health/mental health (additional mental health comments on page 28).

Domain-specific denominator models (DSDM) aren't new. It requires an infrastructure approach to data analytics which needs multi-talented teams to explore almost every cross-section and combination of very large volumes of data, from specific domains, to identify obvious correlations as well as unknown/non-obvious relationships. If there is any doubt about the quality of the raw data, then quality control may mandate data curation. The latter alone, makes the task exponentially complex. Curation may introduce reasonable doubt in evaluating any outcome because the possibility exists that curation algorithms and associated processes were error-prone or untrustworthy (post-curation jitters).

Another demerit for DSDM and the idea of denominator models, in general, may be rooted in the "apples vs oranges" dilemma. Denominator models that underlie science and engineering systems are guided by natural laws, deemed *rational*. The quest for denominator models in data (retail, finance, supply chain, health, agriculture) are influenced, infected and corrupted by irrational[86] human behavior. Rational models of irrational behavior[87] may co-exist elsewhere but remains elusive for data science due to volatility and the vast *spectrum of irrationality* that may be introduced in data by human interference.
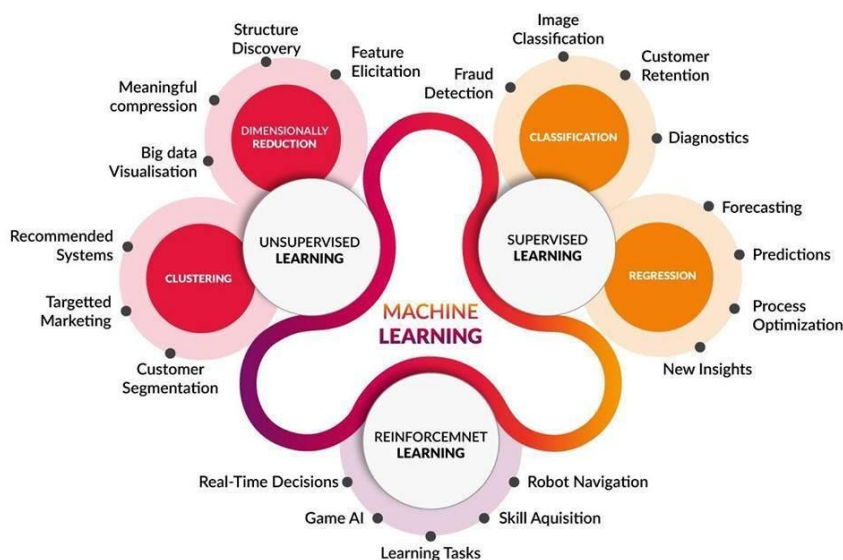

Figure 4 – It appears that we have been mining for patterns and other simpler models (such as clustering, classification, categorization, regression, principal component analysis). But, have we found a set(s) of simpler models or patterns, yet, to test the concept of domain-specific denominator models (DSDM)?

Perhaps the concept of DSDM, ignoring its obvious caveats, may be applied to select domains for specific purposes, for example, healthcare, where deliberate human interference to introduce errors in data is a criminal offense. Case-specific model building, and pattern recognition, may benefit from machine learning (ML) approaches. The latter fueled a plethora of false[88] claims but real success is still a *work in progress* because the bridge between data and decisions will be perpetually *under construction*. Productivity gap and corporate waste are indicators that existing approaches (see Figure 4) are flawed, failing or have[89] failed. We need new roads. The boundary of our thought horizon 'map' is in Figure 4. The tools are incremental variations[90] garnished with gobbledygook alphabet soup. Unable[91] to create any breakthrough, the return of seasonal "winters of AI" indicates the struggle to shed new light in this field since the grand edification[92] during the 1950's. Unable to cope with data challenges, hard facts[93] and difficult progress, the field offered a perfect segue for con artists and hustlers to inculcate falsehoods and deceive[94] the market. Machine learning was substituted[95] by mindless drivel from ephemeral captains of industry and generated hype[96] from corporate[97] marketing machines.

AVOID THIS SPACE – THE DECEPTION SPACE

Data consumers have been led astray by vacuous buzz words manufactured mostly by consulting groups. Part of the productivity gap may be due to fake news, propaganda[98] and glib strategy from smug consultants to coerce large contracts with cryptic "billable hours" to help "monetize" false promises due to "big" data, fabricated[99] claims[100] of "intelligence" in artificial intelligence[101] and deliberately conniving misrepresentations[102] of "blockchain" as a panacea[103] for all problems[104] including basic food safety and security. Callous and myopic funding agencies invested billions in academic[105] industry partnerships to fuel banal R&D efforts orchestrated by corporate collusion[106] and perhaps[107] criminal[108] practices. Abominable predatory practices on display in Africa are disguised under the "smart cities" marketing campaign to mayors of ***African cities, which cannot even provide clean drinking water to its residents.*** Vultures from the industry[109] are selling mayors of African cities surveillance technology and AI in the name of cameras for smart city safety and security. These behemoths are cognizant as to how autocrats use data as an ammunition to plan and justify abuse of its citizens, through algorithms of repression.

EXPLORE THE SOLUTION SPACE – NECESSARY TO ASK QUESTIONS THAT MAY NOT HAVE ANSWERS, YET

Uploading data from nodes along a variety of supply chains is an enormous undertaking given trillions of interconnected processes and billions of nodes with extraordinarily diverse categories of potential data streams, with different security mandates, for example, [a] sensor data about heavy metal (mercury) contamination in water used for irrigation, [b] near real-time respiratory rate of patient with COPD (chronic obstructive pulmonary disease) under remote monitoring telemedicine in rural nursing home, and [c] automated check-out scan data from retail grocery store sales, of fast moving consumer goods, contracted for replenishment (penalty for out of stock) under vendor managed inventory (VMI). The *e-tail* revolution is creative supply chain optimization and reducing retail information asymmetry.

Transforming data and data analytics to inform decision support for *small* cross-sections of examples cited, here, may be theoretically easy in "*power point*" diagrams which "connects" nodes and integrate decisions which circle back to optimize processes, *using pixels*. The reality may be different.

Aggregating data, from various nodes, sub-nodes, devices and processes, on a platform, to enable collective evaluation of dependencies, which could influence outcomes/decisions, may be not only beneficial, but must be mandatory for certain domains, for example, healthcare[110] and clinical[111] environments where patient safety[112] must be of paramount importance.

Any *one* standard platform approach is unlikely to succeed. But a platform approach is probably rational. Multiple segments, on standard (open) platforms, with secure, yet selective interoperable data exchange, between platforms, may address a few challenges. Bringing together various options with respect to data acquisition and analytics, begins to catalyze information flow, decision support and *meaningful* [113] use of data[114]. This suggestion is a few decades old, but not yet a trend, in practice. The drive to connect data was accelerated by the introduction of the concept of the internet of things[115] (IoT). Platform[116] efforts[117] are addressing data[118] upload from devices and sensors[119] (Figure 5).
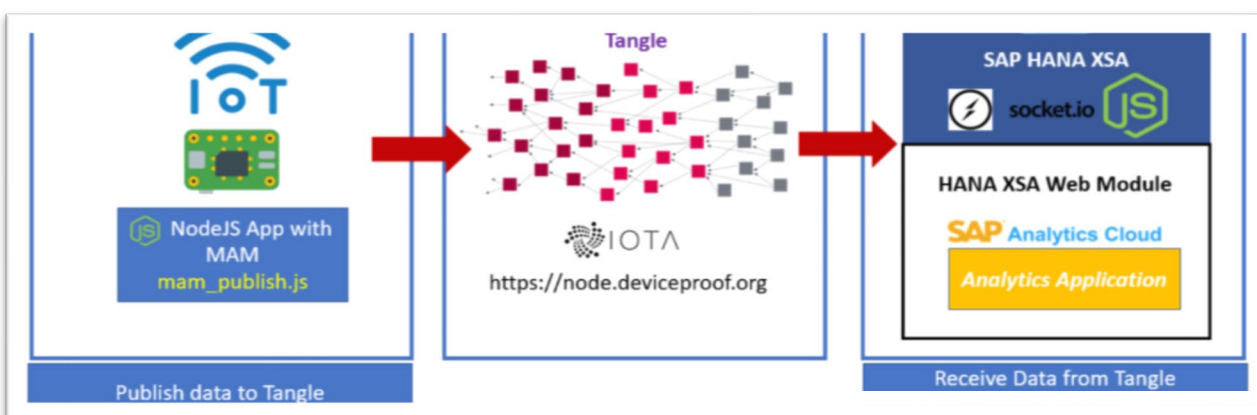
Figure 5 – National and international consortiums, in partnership with large and small software companies, are addressing data acquisition and aggregation. In this example, Tangle appears to be a data aggregation platform (example shows data from a temperature sensor) which can serve data analytics engines to extract information (if data contains information). Replication of "Tangle" for various verticals (retail, health, logistics) and the ability to use open data distribution services[120] may facilitate interoperability between data "holding" services like Tangle. When coupled with supply chain track and trace systems, a retail store (Target, Tesco, Metro, Ahold) can use Tangle data to inform a customer that the One-Touch blood glucose testing strip (healthcare product manufactured by J&J) will arrive at the store (3rd party logistics provider and distribution transportation service) on Monday by 2am and placed on store shelf by 530am (retail store replenishment planning) or delivered to the customer on Tuesday before 9am (online fulfillment services). Any data can be uploaded/downloaded from Tangle.

SOLUTION ECONOMY – WILL WE EVER GET THERE?

There aren't any silver bullets and one shoe doesn't fit[121] all. If we focus on the data to decision process, alone, in any vertical or domain, the variations of analysis and analytics may be astronomical. Initial investments necessary for these endeavors almost guarantees that the extracted value from data (and relevant information) may not be democratized or made functionally available to the less fortunate. In principle, the outcome from data to decisions, when appropriate, must be sufficiently distributed and democratized to provide value for communities under economic constraints. Any ***meaningful*** solution, therefore, is not a scientific or engineering outcome, alone, but must be combined with the economics of technology[122] which must be a catalyst for implementation and adoption by the masses, if transaction costs[123] can be sustained by the community of users.

The economic principle for impoverished environments may be rooted in micro-finance[124] and micro-payments[125] with ***low*** transaction costs (*the downside*: misinformation[126] can be propagated and disseminated at ***low*** cost, too). By eliminating classical "product sales" the focus shifts to delivery of "service" which is a *package* of the product plus other resources (retail mobile banking, infrastructure, telecommunications, cybersecurity[127], security[128], customer service). Users pay (pennies) *only when they use the service*. PAPPU[129] (pay-a-penny-per-use or pay-a-price-per-unit) is a *metaphor* for economic instruments to lower the barrier to entry into markets with billions of users.

The economic incentive for democratization of data is the potential to unleash/create new markets for data, information and decision support, for billions of new consumers (users). The **reward** in the lucrative service economy model depends on harvesting the economies of scale where each user (market of billions) may pay one or more "pennies" (micro-payment for pay-per-use services). The **risk** in the service economy is the collection of that "penny" (per use) at the last step of the *seamless* service delivery process, if the user is satisfied with the quality of service (QoS) metrics. The plethora of partners necessary to create and sustain the ecosystem to deliver the *seamless* service is a herculean task. Sharing a fraction of that "penny" with the partners in the ecosystem is not a trivial challenge. If the QoS delivery metrics suffer due to poor performance of any one partner (component), the end-user "penny" may be unpaid if the QoS metric fails to reach a pre-determined value (time, duration, speed, rate, volume). The inability of one provider (weakest link) in the service supply chain can be financially detrimental to all other supply chain partners due to loss of that penny, *albeit, only for that transaction* (unless the partner has a chronic problem, then, it must be excluded from the ecosystem and the entire value network[130]). Delivery of service is a real-time convergence of operations management which includes (but is not limited to) multiple value chains which must integrate[131] the physical supply chain and the financial supply chain with the service supply chain and customer relationship management (brand expectation).

Determining the cost of execution, to deploy, the example in Figure 5, may be an effective way to study feasibility. By simulating models to explore "what if" scenarios, it may be possible to predict the potential for adoption of services in the context of various economies of scale and PAPPU models which could unchain the economics of technology adoption.

IS THIS FAUX NAÏVETÉ IN ITS PUREST DISTILLATE?

Decision scientists must build a compass to help extract value from data. One compass will not suffice to guide domain-specificity. Existing tools may limp along with *snail-ish* advances (Figure 4) yet it may remain inaccessible to the masses because the tools may not be feasible for mass adoption. The struggle to transform data into information is still in quest of a Renaissance.

The path from *data-informed* to *information-informed* decision remains amorphous. The grand "highway" from information to knowledge is still in the realm of unknown unknowns. Making sense of data is handicapped due to [i] an apparently insurmountable semantic barrier, [ii] scarcity of tools to facilitate location-aware and context-aware discovery of data at the edge or point of use, [iii] lack of standards and interoperability between platforms and mobile devices for data and analytics sharing.

Users in non-OECD nations may not want to idle away while the architects of Renaissance are still in short supply. In the near term, it is necessary that we continue to work on dissemination of data which can deliver at least some value, sooner, rather than later. Decision support based on sensor data analytics may provide economic benefits[132] and incentives, if we can share the digital dividends with the masses, for example, in health[133] and agriculture, including every facet of food, required, daily, globally.

Tangle, a tool[134] to share sensor data using MAM[135] (masked authenticated messaging) may offer hope. Can nano-payments for sensor data address some of the feasibility challenges[136] and pave the way for human-centric economy of things[137] using IoT as a design metaphor? SNAPS[138] is a tiny step in that general direction: distributing low cost tools to enable data-informed decision support for less complex problems. Assuming the Pareto principle is working, perhaps 80% of the problems may be addressed, and even resolved, with simple tools to deliver solutions as a *service*, at the right-time, at the point of use.
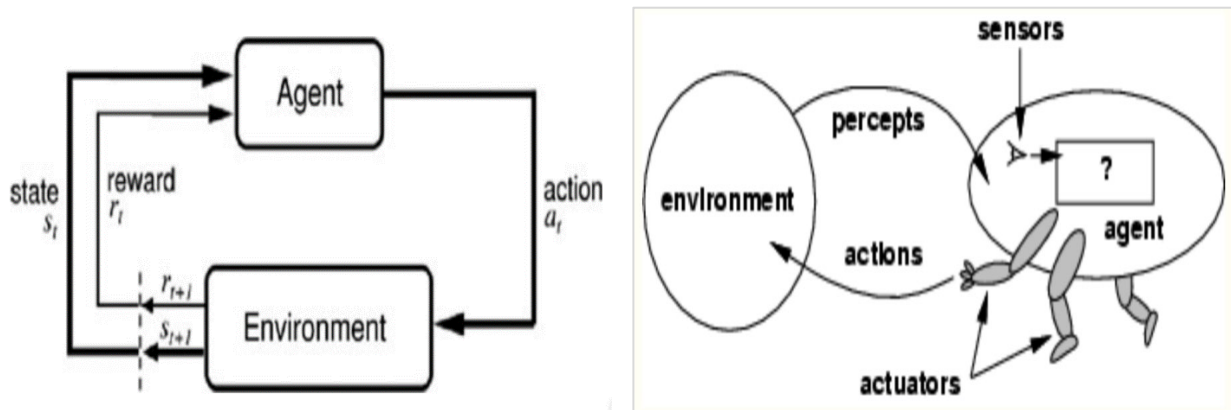
Figure 6 - SNAPS is one layer in a proposed analytics platform (layer cake) which consists of a portfolio of tools aligned with the concept of PEAS[139], a mnemonic borrowed from agent-based systems (**R**), to address systems performance through convergence of percepts, environment, actuators, and sensors. (**L**) Reinforcement Learning[140] (Figure 4), a machine learning technique, compared with PEAS.

REALITY CHECK – DATA FUSION

The inflated view of the sensor-based economy[141] is carefully[142] crafted[143] to create[144] new markets[145] and momentum[146] for sales[147] of sensors and data services[148] aimed to amplify the IoT[149] hype to fortify the deception game. It is promoting the desired effect by spawning mass hysteria and skillfully obfuscating the hard facts which then paves the ground for hordes of consultants to act as "trusted advisors" to make sense of this "revolution" which is supposedly going to change the future of work, life and living. One glaring outcome of delusional[150] propaganda[151] is the near trillion dollar[39] waste related to investment in technology with a failed ROI. Trillions of sensors and devices that *could* connect to the internet (basis for the cosmic scale of IoT) is due to the scale of unique identification[152] made possible by adopting a 128-bit structure in the internet protocol. The unique address spaces in IPv6[153] is **29 orders of magnitude higher** than IPv4 if one compares[154] 4.3×10^9 address spaces for the 64-bit IPv4 versus 3.4×10^{38} unique address spaces for the 128-bit IPv6. New possibilities[155] and applications[156] may arise due to the flexibility of IPv6 to directly connect to the internet (rather than sub-nesting under/via gateway nodes).

The difference between promise and perils in deploying the concept of IoT as a design metaphor is rooted in grasping the difference between connectivity, discovery and actionable insight. Just because something is *connected* does not mean value emerges, automatically, without a **connected ecosystem**. If a visitor's tablet can discover the printer in an office and use it to print a meeting agenda, then we have extracted some type of value between the connectivity of the tablet and the printer, which were able to "discover" each other, and that discovery enabled the gain in efficiency (printing the agenda). In its basic form, this is an example of very simple data fusion which leads to an actionable output and provides "information" for the meeting attendees in terms of the printed agenda. Connecting trillions of entities to the internet is an exercise in futility unless discovery and data fusion enables meaningful extraction of data to move up the DIKW[157] value chain[158] where data precedes information, knowledge and wisdom.

Figure 7 - PEAS, a mnemonic borrowed from agent-based systems (**R**), addresses systems performance through convergence of percepts, environment, actuators, sensors. The OODA[159] loop (**L**) and PEAS[133] contribute to advance DIKW (data, information, knowledge, wisdom), which begins with data fusion[160].

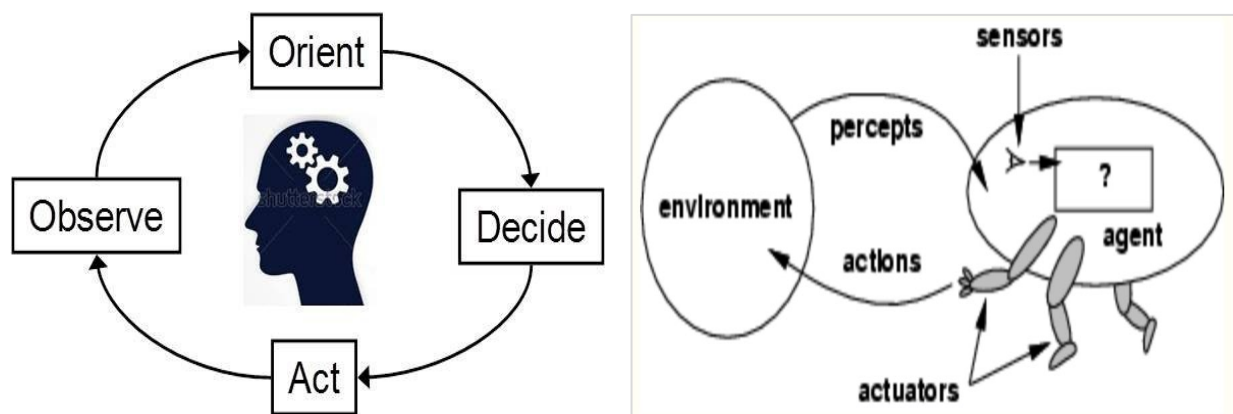The PEAS paradigm resembles OODA (Figure 7) because "observations" refer to scanning (sensing) the environment and "orientation" informs the image of the environment by encapsulating both descriptive and predictive analytics ("decide" includes prescriptive analytics). The integration of data fusion and analytics with agent-based systems is critical in the era of IoT. The networked society faces a deluge[161] of data yet the human ability to deal with data, analytics, and synthesis of information may be inefficient. How can devices discover data and facilitate processes without intervention by humans? Automated on/off action taken by a domestic thermostat and HVAC based on temperature sensors may be quite primitive when considering autonomous objects in air (UAV), land and water.

Raw sensor data unless discovered and combined with "perceptions" from the environment, may be context-deprived and over/under utilized, which lowers the value of the data with respect to the desired goals. The perception from the environment is not unique but a "learning" task for the system. It may re-use the experience (learning), when relevant and appropriate, at a different instance (Figure 6). Can this "learning" become mobility-enabled and "teach" other devices, for example, by transmitting a *tutor* virion to another computer or drive or system? Can this device communicate in natural language and/or respond/understand the semantics in human queries?

Taken together, unleashing the value of data may require coordination of ABS (agent-based systems) in every facet of our interaction with machines, objects, and processes which may benefit from feedback. ABS is an old[162] concept[163] but resistant to succinct definition[164] because agent activity must remain agile and adapt to the operating objective (PEAS) and problem context (OODA). Equation-based models create rigid, hard-coded software. Agent-based system design induces agility, may enable "drag & drop" variant configuration to adjust (on-demand) to volatility, uncertainty and ambiguity, inherent in most environments. In the context of democratization of data and benefits for the masses, agents can be highly personalized and "belong" to people, for example, personal agents, as discussed[165] elsewhere, with respect to cybersecurity. A similar modus operandi can be adopted for other use cases where data fusion[166] can be dynamic and composable (composed when necessary, depends on context) not only for use-cases but also for individual user-specific cases/applications.

Agent Characteristics	Definition
Autonomy	Operates without the direct intervention of humans or others
Sociability	Interacts with other agents, that is, communicates with external environment such as sensors, fusion systems and human operators
Reactivity	Perceives its environment and responds in a timely fashion
Pro-activity	Exhibits goal-directed behavior by taking the initiative
Learnability	Learns from the environment over time to adjust knowledge and beliefs
Mobility	Moves with code to a node where data resides
Anthromorphicity	Externally behaves like human

Table 1 – Generally, Agents are computational entities (software) designed to perform specific tasks, autonomously. Agents embedded in devices (sensors) may have logic capabilities to perform artificial reasoning tasks (ART) and/or optimization[167] in multi-agent systems (MAS)

The role of software agents to "discover" and then determine which data and/or data fusion may be meaningful or relevant (user-specific), is an old idea, waiting to be effectively applied. Connecting data must be contextual. The *established* contextual relationship must be discovered and "understood" by agents or group of agents. Another old idea is to *pre-establish* the context based on knowledge graphs. The common thinking that W3C standard RDF[168] (resource description framework) triples are the solution for knowledge graphs is incomplete. This myopia, is, in part, one reason why the semantic web[169] failed to flourish. The brilliant idea of representing subject-predicate-object (SPO) as a relational RDF graph is certainly useful and applicable in many instances but the approach *bites the dust* when the reductionist 1:1 granular relationship fails to represent reality. The latter is painfully obvious, especially in medicine and healthcare, where the rigidity of the RDF standard structure and RDF schema may be an anathema. The "force-fitting" of RDF to healthcare applications[170] oversimplifies scenarios to the point where it may, inadvertently, introduce errors, simply due to exclusion, which may prove to be fatal.

One proposal suggests adapting[171] RDF by creating relationships between sets/subsets (rather than points and vertices as in classical SPO) using the set theoretic[172] approach. It is easy to grasp why "set" of symptoms and potential set of causes may make more sense in medicine and healthcare. The overlapping (Venn diagram) subset of relationships may be indicative of likely causes for symptoms. Generic symptoms, for example, fever, can be due to a plethora of causes and why a rigid 1:1 relationship in RDF could turn lethal in healthcare applications. The finer granularity of RDF is a disadvantage yet it is key to merging attribute lists about an entity sourced from different data sources. The latter enables better search and discovery across diverse domains, the hallmark of globalization of enterprise systems.

An even older idea[173] which is recently[174] enjoying scientific[175] as well as public attention[176] is the labeled property graph (LPG). It is suitable for use cases which may be focused on providing stores for single applications and single organizations, such as, domain-specific denominator models (DSDM). LPG proponents are less committed to standardization, interoperability and sharing. It is in contrast to the W3C ethos and RDF which favors standardization, interoperability and sharing, which makes it useful in discovery using graph pattern searches. Optimization of local (domain specific) searches[177] using graph-traversal algorithms[178] are better suited for property graph (PG) databases. Knowledge graph networks embedded between sets/subsets may give rise to amorphous "linked" clouds, which could be industry[179] specific and may be domain specific[180] as well as user specific. Imagine if data from each patient could be used by an ***automated knowledge graph engine*** to create precision, patient-specific and personalized knowledge graphs. Extracting the relationships and contextualizing the relevance of symptoms may influence the accuracy of diagnosis. When viewing knowledge graphs in a population study (epidemiology), it may be easier to detect outlier events or cases that did not fit expected patterns.

Therefore, domain specific denominator models (DSDM) may be represented as a domain/user specific knowledge graph networks. Agents may be invaluable in working within this environment to discover relationships and contexts (specificity reduces search space), as well as discover data sources, and perhaps, based on embedded logic, decide whether the features or attributes calls for data fusion.

For any agent-based approach to succeed, it is critical that the agent framework and standards are interoperable with the knowledge graph network and the data domains where the agent is searching. The opposing tendencies of RDF vs LPG in terms of standardization, interoperability and sharing may limit agent mediated "cross-investigation" of domains, discovery and data. Therefore, it begs to question the expectation why one agent must perform in all domains. Perhaps, the success of agent search and discovery depends on semantically annotated structured data. The latter depends on ontological structure. W3C proposed[181] OWL standard web ontology language [182] and recent variations (VOWL[183]) may contribute to interoperability. The old idea of Internationalized Resource Identifier (IRI), as a complement to the Uniform Resource Identifier (URI)[184] to identify resources (to facilitate discovery) is a valid principle but yet to be adopted in practice. The plethora of old ideas (referred here) suggests that the value of these ideas may have to be revisited. This field needs new "blood" and new "eyes" to imagine new ways to address interoperability. But, in reality, today, on top of this wobbly incompatible infrastructure, we are layering the "snake oil AI" and unleashing an incorrigible torrent of half-truths.
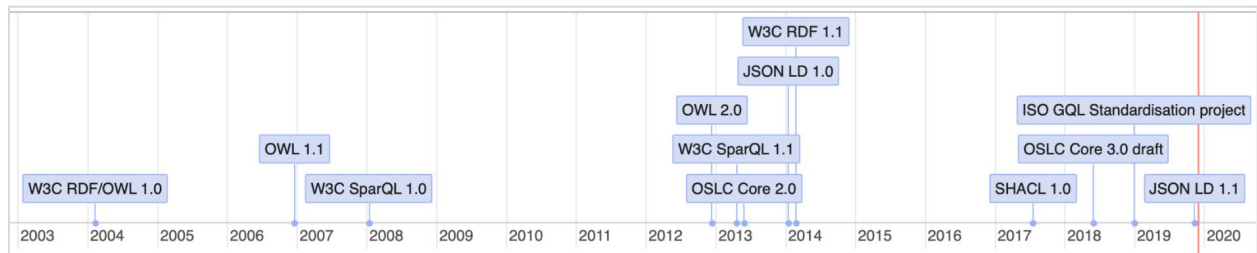
Timeline 1: Long march of graph related[185] specifications. Is standardization a sufficient solution?
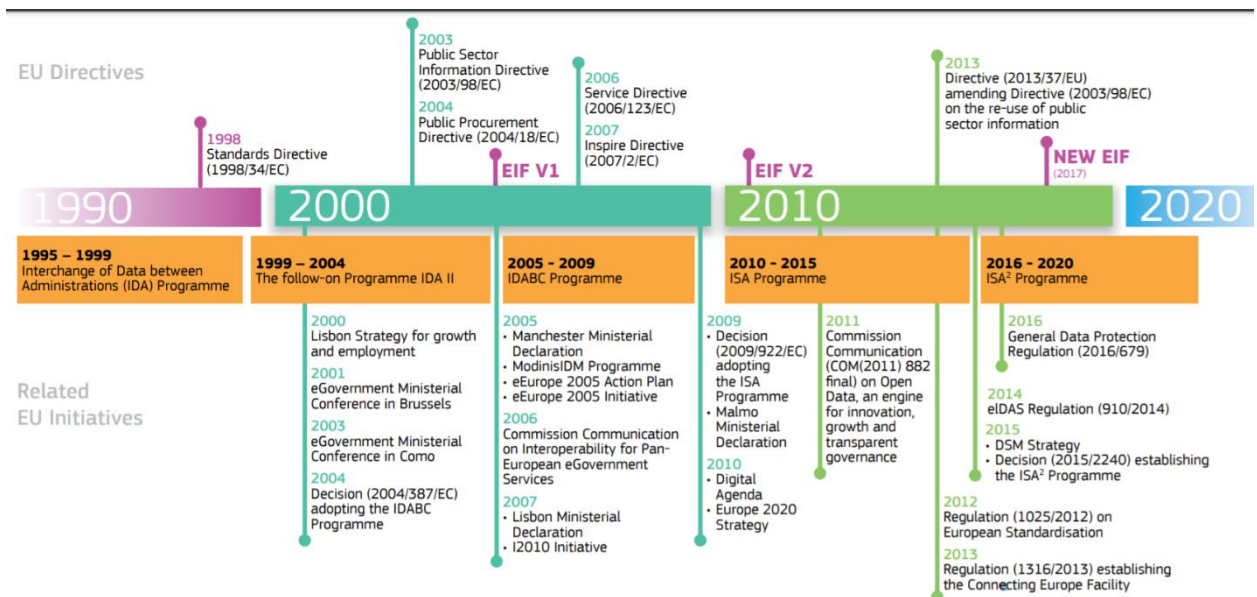
Timeline 2: EU's Elusive[186] Quest for Interoperability: Is 30 years not enough?

"DOUBLE A" PERSPECTIVE OF DATA AND TOOLS vs THE HYPOTHETICAL POROUS PARETO (80/20) PARTITION

If combined, Africa (1+ billion) and Asia (4+ billion) approximates 75% of the world's current (2019-2020) population. The corporate view of 75% of the world's market is driven by the promise of new markets, new customers and new wave of consumerism. It has little to do with lifting the lives of people. Our discussion about the physics and mathematics of data, therefore, is *a tempest in a tea cup*. For 5+ billion people, the trial and tribulations of data and data analytics, we have discussed here, can be dismissed with an eye roll. It is useless for most pressing daily applications in the "double A" world.

Thus far, what we have discussed, on one hand, may be an exploration of the tessellated facets in our search for meaning, and on the other hand, it is a discussion which may find parallels with the "six blind men and the elephant" syndrome[187] apparently divorced from complementarity[188] or synergy. It is as if the outcome from 20% of the global population, relevant or not, is thrust upon the remainder of the world market. In 80% of the cases for 80% of the global population the daily decisions about FEWSH (bare necessities of life: food, energy, water, sanitation, healthcare) do not require artificial intelligence, machine learning algorithms or optimization of 'state space' for hundreds of variables.
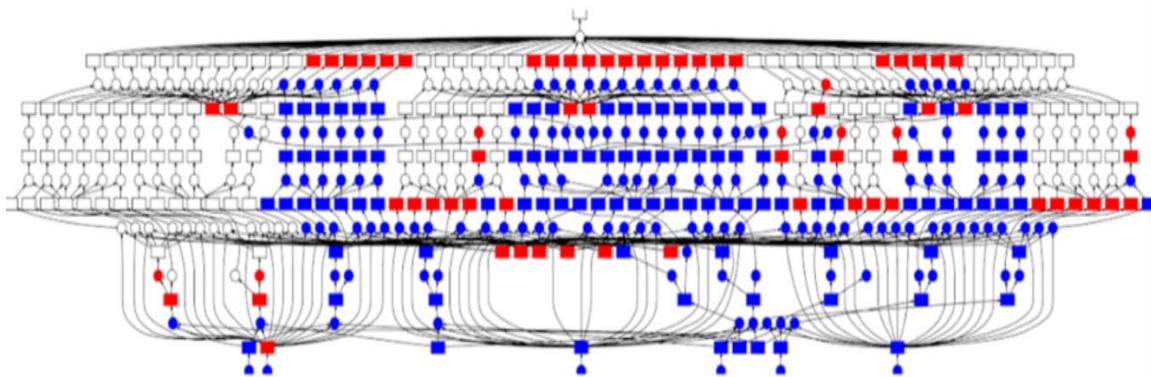
Figure 8 - Blind men and elephant. Each man guesses his own part of the elephant but blinded by hype[189] they cannot perceive the "whole" elephant. A metaphor for focus on parts, which occludes the system. Cartoon (bottom): penchant for decision trees by *power-point* rather than search for low hanging fruits.

In 80% of the cases for 80% of the global population the daily decisions about FEWSH require data, *small data*, data in near real time and data that only impacts and enhances the user. In that context, jumping up and down about *democratization of data* is tantamount to chest-thumping. The data in these use cases may be related to a subset of FEWSH (food, farm, agriculture, water, healthcare). If the tools are there to acquire this data, then the data is available. Therefore, is democratization really an issue? Is it a politically correct word that the 20% world prefers to use as a hand-waving advocacy of problems that are divorced from reality on the ground? Is "democratization" a "theme" song for advocacy groups in OECD nations who may display the symptoms of the *six blind men and elephant* syndrome?
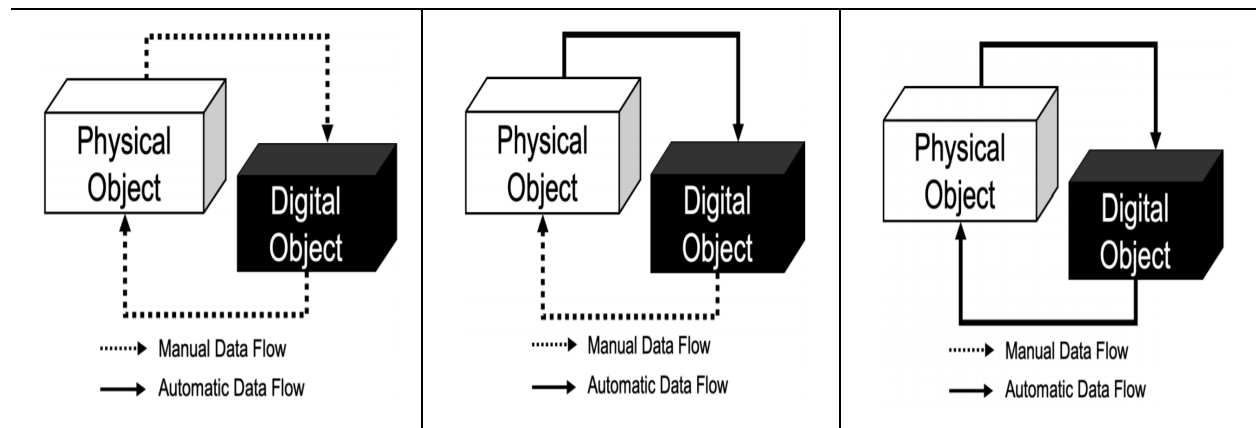
Figure 9 – Digital Duplicate (left), Digital Shadow[190] (center), Digital Proxy (right) and Digital Twin (bottom) are variations of digital models of physical objects, integrated with data flow. But, do we know if it is *meaningful* for data related needs for 80% of the world? It is unlikely to be solved by Digital Twins[191] or flamboyant gimmicks peddled by fake pundits on the pages[192] of *Forbes.* However, the R&D related to these tools may trickle through the "pores" from the 20% side of the partition to the other side (80%) of the *porous* Pareto[193] partition and occasionally[194] may be helpful.

The number and provision of central DXA units available in the EU27 (Data on reimbursement and waiting time [10])

	DXA units/million	Waiting time (d)	Cost (€)	Reimbursement
Austria	28.7	14	30 [11]	yes
Belgium	53	14	34 [12]	partial
Bulgaria	1.2	0	59 [9]	none
Cyprus	23.9	20	75 [9]	yes (depending on income)
Czech Republic	5.2	40*	32 [9]	yes
Denmark	14.6	30	187 [13]	yes
Estonia	8.9	14	14 [14]	yes
Finland	16.8	1	146 [15]	yes
France	29.1	14	41 [9]	yes (conditional)
Germany	21.1	0	36 [9]	yes
Greece	37.5	11*	115 [9]	yes
Hungary	6.0	15*	7 [16]	yes
Ireland	10.0	140*ª	99 [17]	yes (conditional)
Italy	18.6	83*	81 [9]	yes (conditional)
Latvia	4.9	10*	18 [18]	yes
Lithuania	3.4	6*	28 [19]	no
Luxembourg	2.0	30	59 [20]	yes
Malta	9.7	105*	184 [21]	yes
Netherlands	10.7	14*	84 [22]	yes
Poland	4.3	1	10 [9]	yes (conditional)
Portugal	26.9	8	5 [23]	yes
Romania	2.4	7	5[24]	yes
Slovakia	10.7	18*	32 [9]	yes
Slovenia	27.1	11*	29 [25]	yes (conditional)
Spain	8.4	105*	109 [9]	yes
Sweden	10.0	60	152 [26]	yes
UK	8.2	11*	51 [27]	yes[d]

Table 2 – Availability[195] of DEXA (dual energy x-ray absorptiometry) scan machines to measure bone mineral density (BMD), a fair prognosticator for osteoporosis. The European standard[196] is 11 (DEXA) DXA units/million. In an updated estimate, the poorest country in EU27 offers 4 machines/million[197] whereas Bulgaria's neighbor Greece boasts 37.5 DXA units/million. In comparison, Indonesia[198] has only 0.13, India[199] 0.18 and Morocco[200] 0.6 DXA units/million. For the health of the people in these nations, how can democratization of data lower their risk of osteoporosis? Are we asking the correct questions? Are we pursuing the wrong reasons? Are we arm-chair analysts helping the (BMD) medical device industry[201] accelerate their sales campaigns to AA nations? Can data provide relevant answers?

Table 2 offers a glimpse of one problem in healthcare. The tools to acquire the data are in short supply. Measuring the risk of osteoporosis is a prerequisite for prevention and treatment, if affordable. Arm-chair "scenarios" of medical IoT will want to connect DEXA (DXA) scan data with sales of milk and exposure to sunlight as a "wellness" indicator. From the "double A" perspective, it may be a futile "power point" exercise because milk *may not be available* for the age group[202] generally at high risk of osteoporosis in the AA nations. In most parts of Africa and Asia there is an opulence of sunlight.

Just because there is an "IoT" scenario, does not mean it is worthwhile or valid for users in "double A" nations. Just because there is data also does not result in information. Can we reduce incidence rates of osteoporosis simply by adding more DXA machines per capita? A recent (2013) study using seven national electronic healthcare records (EHR) databases revealed that Denmark (14.2 DXA units per million) showed age- and sex-standardized incidence rates (IRs) of hip/ femur fractures 2X higher than those observed in UK (8.2 DXA units/million), Netherlands (10.7 DXA units/million), and Spain (8.4 DXA units/million), while Germany (21.1 DXA units/million) yielded IRs in the middle range.

DRUG	PRESCRIBED FOR	UK PRICE	US PRICE	PRICE RISE
NEXIUM per 20mg tablet	Acid reflux	£0.66	£7.40	1120%
ACTIMMUNE, 12 vials	Genetic diseases, osteopetrosis	£5,400	£42,990	800%
DARAPRIM per tab	HIV, cancer, malaria patients	£2.30	£619	26,900%
NASONEX, 50mg	Nasal allergies	£7.68	£224	2,900%
CINRYZE, 2 vials	HAE, genetic disorder	£1,336	£3,645	200%
HARVONI per tab	Hepatitis C virus	£464	£928	270%
SOVALDI per 400mg tab	Hep C in children under 12	£416	£855	200%
DIAZEPAM, per tab	anxiety, relaxation, muscle spasms	£0.02	£3.05	15,200%
OVEX, 100mg tablet	Threadworm parasite	£2.54	£300	11,800%
LIPITOR, per 10mg tab	Statin	£0.46	£4.50	980%
VIAGRA, per 25mg tab	Male impotency	£4	£61	1,500%
ZOCOR per 10mg tab	Statin	£0.64	£4.20	656%
CYMBALTA per 30mg capsule	Anti-depressant	£0.80	£9.48	1,200%
EPIPEN, 300mg	Allergies	£52.90	£523	1000%
HUMALOG INSULIN	Diabetes	£16.61	£215.30	1300%
HIP REPLACEMENT OPERATION		£7,313	£26k–£37k	350%
KNEE REPLACEMENT OPERATION		£6,315	£24,801	390%
CATARACT OPERATION		£803	£5,780	720%

Table 3 – Plague of unethical profitability makes US pharmaceutical[203] business model in healthcare an abomination which is inappropriate for mimicry in any part of the world. Source: Dr James Nolan[204]

CONUNDRUMS

On one extreme we have presented sophisticated ideas for making sense of data. On the other hand, we doubt whether the toothless call for *democratization of data* from the affluent 20% of the world can help to lift the lives of people on the other side of the porous Pareto partition (80% of the world). It is not a true "Pareto" scenario, but the 80/20 nature of this problem evokes the Pareto principle as an analogy, hence, Pareto partition. The R&D outcome of the 20% may contribute certain elements to the 80% side. The "partition" is a metaphorical porous membrane, with bidirectional porosity, if necessary.

Often, it may not be necessary. For example, is it necessary to deal with data and data models in this scholastic[205] manner (Figure 10) for all problems? The 80/20 global partition may be prominent in agriculture, healthcare and energy. In case of the latter, what is the value of smart metering or load balancing algorithms when there isn't enough energy, at an affordable cost, to supply the basic tenets of economic growth? How many farmers in the "80%" world can afford to use the drone-on-demand[206] system? Why should people from the majority sector (80%) need useless marketing tools[207] when daily healthcare for the less fortunate can solve a myriad of problems with *just-in-time little bits of data*, for example, daily blood glucose level from a diabetic (versus the *always-on* real-time monitoring of blood glucose) or monitoring individuals for silent myocardial ischemia[208], a leading contributor to death. This is the debate where the economics of technology and its *relevance to the community* are crucial issues which may **enable** adoption or **disable** the dissemination of technology, which could have contributed to economic growth, workforce development and sustainable job creation.
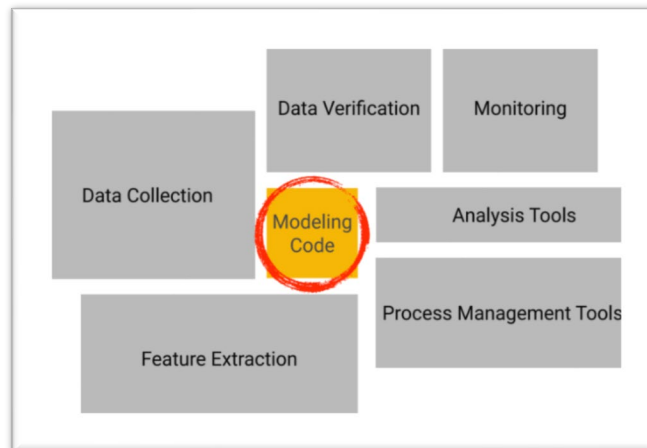
Figure 10 – From monitoring an event to using the data to inform a decision, there are a plethora of steps[209] in the standard operating procedure (SOP) for the "20%" deploying data to drive decisions. However, irrespective of socio-economic issues, in future, all aspects of feature selection and feature engineering may emerge as a pivotal or rate limiting step in dealing with diverse data sources. In this context, automated feature extraction and other feature related steps may be a very significant step. *In combination, automated feature engineering and automated knowledge graph engines (see page 16) may usher new dimensions in data and data analytics, if automated data curation could improve data quality.*
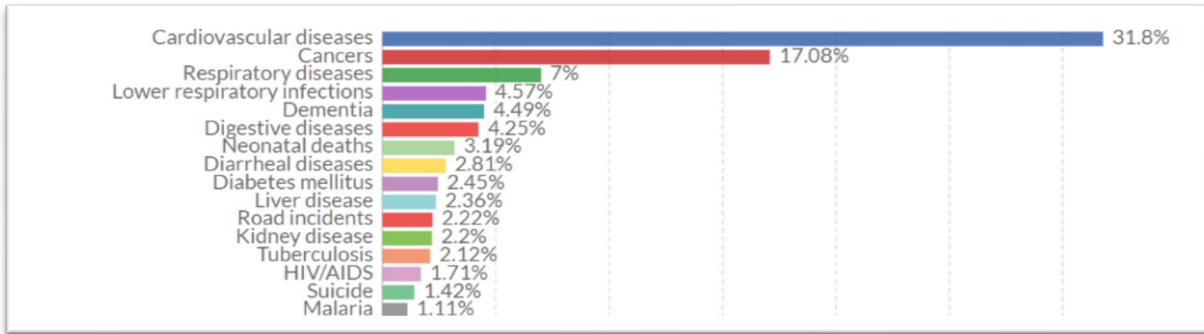
Figure 11 – Share of deaths, by cause (2017) percent of total deaths[210]. Data refers to specific cause of death, which is distinguished from risk[211] factors for death (water and air pollution, diet, sanitation).
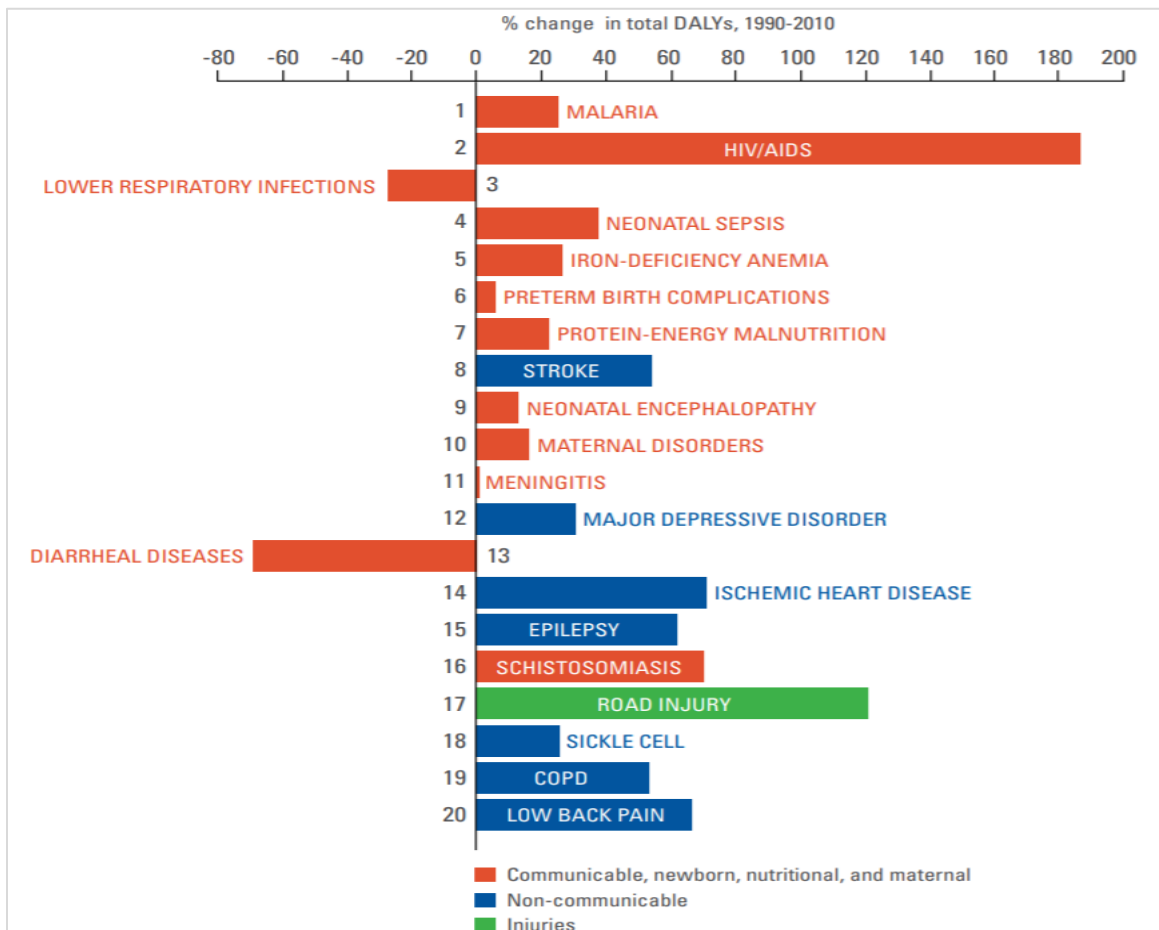
Figure 12 - Shifts[212] in leading causes of DALYs for females, Ghana (1990-2010). The leading 20 causes of DALYs are ranked from top to bottom in order of the number of DALYs they contributed in 2010. DALYs (Disability-adjusted life years): The sum of years lost due to premature death (YLLs, Years of life lost due to premature mortality) and years lived with disability (YLDs, Years of life lived with any short-term or long-term health loss causing disability.). DALYs are also defined as years of healthy life lost.

Worldwide, Africa[213] accounts for 9 out of every 10 child deaths due to malaria, for 9 out of every 10 child deaths due to AIDS, and for half of the world's child deaths due to diarrhoeal disease and pneumonia. More than one billion children are severely deprived of at least one of the essential goods and services they require to survive, grow and develop[214] - these include nutrition, water, sanitation facilities, access to basic health-care services, adequate shelter, education and information. As a result, almost 9.2 million children under-five die every year and 3.3 million babies are stillborn. Most of the 25,000 children under five that die each day are concentrated in the world's poorest countries in sub-Saharan Africa and South Asia. There, the child mortality rate is 29 times greater than in industrialized countries: 175 deaths per 1000 children compared with 6 per 1000 in industrialized countries.

These facts (paragraph above) and Figure 12 offers a vastly contrasting view to that of data tools and democratization of data as essential for lifting the lives of the people living on the majority side of the porous Pareto partition. Simple forms of small amount of data, sufficiently informing ordinary tasks, may be suitable for delivery of global public goods and services to the majority of the 80% world. It is absolutely ludicrous to think that "big" data, AI/ML, blockchain or smarmy publicity[215] stunts may help, in this context. What we need is the concept[216] of "bit dribbling" perhaps coupled with pay-a-penny-per-use (PAPPU) systems to help people improve their quality of life without the constant need for charity.

Technology may play a central role to reach the billions who need service but not in the form of business[217] which is staple in the West and copied by the thoughtless Eastern schools, especially in India. Technical tools will generate data. The ability to use that data, judiciously, may be key to the value of data, for impoverished nations. Coupling social need with technical catalysts must be optimized in the context of the community and not according to *Wired* or *MIT Tech Review* or *HBR*. Advanced R&D is the bread and butter of progress, but the application of advanced tools must be contextual to the services that the community can sustain. Just because auto maker Koenigsegg claims the *Agera* model was built with a "less is more" philosophy does not mean it is the pragmatic standard of transportation suitable for Calcutta, India. In the realm of systems engineering courses and education, dynamic optimization (DO, Figure 13) illustrates a similar perspective. The principle is worth teaching, worldwide, but the practice must be relevant to the case. Do we all need DO in everyday life and living? Is it necessary for all types of edge analytics to process data using convolutional neural networks (CNN) on a mobile device or phone?

The conundrum of *not* applying the tools we think we have mastered is counterintuitive to the problem-solving ethos in the 20% world. We are ever ready to use the latest and greatest knowledge from the bleeding edge to derive and drive the best possible 99.99% perfection and performance. The quagmire of lies aside, we do have real tools which offers notable advantages. But, the volume of the 80% of the world and the economic handicap in these communities must be assimilated in order to change our thinking. The acronym KPI (key performance indicator) is for "performance" which is euphemistic for profit in the 20% world. It may or may not be in the best interest of the people. For 80% of the world perhaps KPI should stand for "key people indicator" and ascertain whether we improved lives of people.

Improving lives, however, is relative to the life you aim to improve, a life with disabilities[218] versus life with social[219] void are active domains in robotics. Indeed, robotics is key for the 21st century but the robot propaganda, written mostly by hacks[220] and driven by media[221] sales, is a sign of the times. Essential robotics and robots for tasks that are dangerous, dirty and dull (repetitive) are progressing in various domains.

The idea of the automated robotic factory was popularized by Philip K. Dick's fiction "Autofac" published[222] in 1955 (*Galaxy* magazine). The "lights-out" automated manufacturing facility FANUC[223] (factor automated numerical control) is in operation since 2001, in Japan. But FANUC is an exception. Even though "lights-out" robotics made significant strides in heavy industry, it is far from the Orwellian scenarios promoted through chicanery[224] and buffoonery[225] aimed for profit from book sales by discombobulating the masses. In reality, the promise of robotics must be balanced with the degree of **trust[226] in automated execution**, especially when humans are involved, directly.

In other instances where human life is at risk (for example, transportation, manufacturing, mining) the trust in automated action (robot) is as good as the planning for "what ifs" **when** the auto execution goes awry. But, that is a deterministic perspective where what could go wrong is anticipated, albeit with some degree of uncertainty. However, if a mobile robot crashes with a holonic manufacturing podium, it may generate a cascade of events where the outcome may be non-deterministic. The critical question in such a scenario is the extent to which a non-deterministic outcome can be tolerated and the *acceptable* cost of risk despite the 'open-ended' uncertainty. Few can even approach to answer this/these questions[227] because it verges on the domain of unknown unknowns.

But, that may not deter simulation aficionados from pursuing stochastic (*what if*) models to capture distribution of randomness in non-deterministic outcomes. Heuristics approaches may surface to suggest contingency measures. This is "video gaming" of automation[228] which could turn deadly in reality. The *executive* robot may be suited for "3D" tasks (dull, dirty, dangerous) but unsuitable for relinquishing human oversight and control if lives are at risk. However, even worse are evil acts perpetrated by humans to bury[229] and ignore[230] the failure of automation, in the pursuit of profit.

Robotic tools in the 20% world are engaged in sophisticated activity which may be subjected to high oversight. In general, the 80% world is not a customer for such implementations in terms of mass consumption. Automation replacing or reorganizing jobs is not a new event (for example, auto industry) because technology[231] shifts the cycle of jobs and with it, the economy. Rapid changes in skill sets and the volatility of job categories influence other domains, namely, K-16 education, training, skills development (capital, labor market, employment) and communication (hopefully, the truthful variety). The rate of change in certain ecosystems are dreadfully slow (for example, education system) whereas the evolution of the job market may resemble the rapid pace of bacterial growth, albeit slower than viral growth rates. The diffusion of robotics will take time and only if the building blocks of automation can be popularized, globally, in a manner that Lego blocks may have inspired young minds to compose, create and construct.

The 80% world can benefit from robotics, for example, by reducing global disease[232] burden in emerging economies and developing nations. If sewer system cleaners (mostly young children who are flexible enough to reach cramped spaces) are replaced by robotic tools to do the "dirty" work, then the health risk to the young workers and public health risk to the community (from unsafe or unhygienic sewers) may be reduced. One must balance between improving lives versus affordable robots to perform such tasks. Robotics (tools) are an ally for the underserved by helping to improve lives, enabling health and wellness, rather than fear-mongering and flagrant deception[233] how robots will replace human jobs.
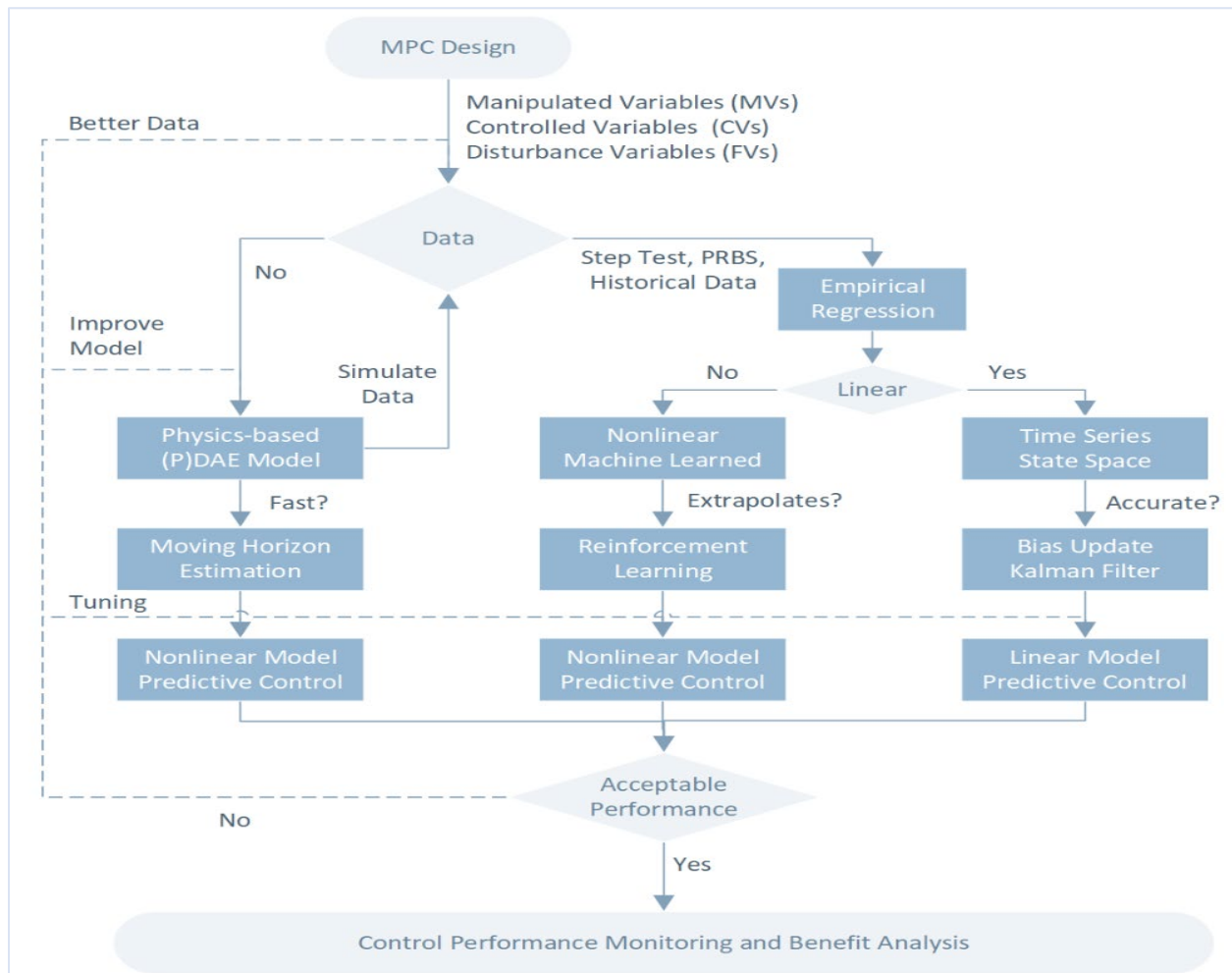
Figure 13 – Dynamic Optimization[234] is a central component of systems engineering where applications of numerical methods for solution of time-varying systems are used to improve performance and precision of engineering design and real-time control applications, which may have a broad spectrum of use, for example, from optimizing the artificial pancreas to fuel cells. Principles of DO may be taught[235] worldwide but DO, systems level data science[236] and Bayesian[237] statistics are excellent tools yet, often, less useful for 80% of the tasks for 80% of the world on the other side of the porous Pareto partition.

REFLECTION

The protagonists of the "open source" movement may be dismayed by "partition" suggestion which does not disguise an "ours" vs "theirs" view of a divided world. It is unfortunate but necessary to serve as a constant thorn in our conscience and sow discomfort. The "partition" thinking originates from the corporate pursuit of developing a smörgåsbord of bleeding edge tools in the 20% world and then coerce 80% of the world to buy such products and services ("next billion users"). To add insult to injury, corporations from the 80% world are salivating to acquire rights to these products and bring it to their market (for example, Tata (TCS), TechMahindra, Wipro, Infosys, "body-shops" in India[238]). There exists a *nano-cosm* of people in the 80% world who could be a part of the 20% world. Because they are an influential minority and holds the financial power in the 80% world, they are aligning their "profit" objectives with the 20% world.

This mismatch may be at the heart of this global dilemma and creates the necessity to consider the porous Pareto partition in terms of people and service for the end-user. People in AA nations are not the consumers for facial recognition software systems. The abuse[239] is perpetrated by governments. People in the 80% world are not seeking quantum computing to process exabytes of data. People are seeking simple information, for example, for their health (data for blood cholesterol level) or from their farm (data about concentration of heavy metal contaminants in irrigation water used for fresh produce, such as, tomatoes). These services help people, the end user, the consumer. This discussion is about what science, engineering and technology businesses can do for people where the key performance indicator is user-centricity and human-centric[240] well-being.

This mismatch between the business to consumer (B2C) services versus the business to business (B2B) services is not new. The 80% world is always looking to the 20% world when planning strategic moves for climbing "up" the supply chain. The fact that the tools from the 80% of the world may not fit in the 20% world is obvious from "frugal innovation" calls[241] by others. Yet, the type of imagination, invention[242] and innovation[243] from the 20% world R&D is quintessential for all and a few are immensely helpful in the lives of the 80% world. What may be often lost, is the *translation* of the advances from the 20% world, for people-centric applications, in the 80% world. This discussion is not singing the praises about the investment in research that only the 20% of the world can afford to push forward because we know[244] the facts. The world is indebted for the strides made possible due to entrepreneurial innovation in such havens such as Massachusetts and California. This discussion is about exposing the lies[245] but not slowing the leaps of vision from the 20% world which may help create low cost[246] tools to serve the 80% world. The world needs R&D from the 20% world, hence, "porous" partition may facilitate the flow of innovation. Not "as is" but with contextual modifications to better serve communities in the 80% world, at a self-sustaining cost (for example, the PAPPU model, pay-a-penny-per-use or pay-a-price-per-unit).

Mental health is one problem where "porosity" is most welcome because most of the world are affected by generic[247] as well as specific issues, which contributes to economic[248] drain. Inherited bipolar and unipolar disorders[249] do not discriminate on the basis of race, color, religion or national origin. The neurochemical, neuroendocrine and autonomic abnormalities associated with these disorders need biomedical research to elucidate the neurobiological basis of these diseases. The latter is not feasible for the 80% world. Harvesting data[250] from external symptoms and pattern analysis[251] may offer a low-cost substitute, to inform the nature of treatment required. People in the 80% world may find it useful.

However, this discussion is not a *to-do* list. It is not a roadmap. It may be a compass, oscillating asynchronously from esoteric thoughts to bare necessities. We are immersed in this duality. One cannot exist without the other. The role of the "partition" is to help focus on the issues that are unique to the environment and community that we wish to serve. It is not a partition of R&D or people or products but a partition for *delivery of service*.

The idea of democratization of data is a bit buzzy but gimmicks are key to marketing. August institutions, including MIT, are complicit in sponsoring potentially puffy pieces to keep the hype[252] alive. But, the fact remains that enabling data to inform decision is a bedrock of measurement, central to all, irrespective of economic status. The porous Pareto partition is a catalyst to focus on services for the 80% world, where less could be more and serves our sense of égalité.

Dribbling a bit of data to inform a person that her respiratory rate (RR) is fluctuating, too often, may be a preventative measure (think of the proverb *a stich in time saves nine*). Informing the person that her RR data is not copasetic, may reduce future morbidity due to COPD (chronic obstructive pulmonary disease). Providing data and information may be *without* impact on the quality of life in the absence of follow-up (clinic). In terms of data and information, alone, by enabling something simple and even mundane, the people-centric application of technology and data, preferably at the edge[253] (point of use), may help to do more, with less. Unbeknownst to us, we are attempting to use the pillars of science, engineering, data, information and knowledge to build bridges which may serve as a platform to provide service to billions of users. Rather than *gilding the lily* we are offering a "bare bones" bridge which serves a rudimentary purpose and still may exclude a few. The volume and demand for such low cost services [pay-a-penny-per-use (**PAPPU**) service business] may also be profitable for the business ecosystem.

Supporting a sustainable effort, to lift the quality of life, will depend on the extent of the product *ecosystem* and many other "things" in addition to technical and sensor data (see "disclaimer") as well as the *cohesion of the service supply chain*. Socio-economic data[254] and related factors are equally significant. Core elements are education of women[255] and trust[256] in women, followed by civic honesty[257], social value[258] as well as inculcating the practice of ethical profitability in social business and entrepreneurial innovation to accelerate the pace of creating pragmatic tools and solutions for remediable[259] injustices.

DISCLAIMER – THE ILLUSION OF DATA, DELUSION OF BIG DATA AND THE ABSENCE OF INTELLIGENCE IN AI

Neither data nor AI[260] is a panacea. The acquisition of data and analysis of data is not a guarantee that there is information in the data or that the information is actionable in terms of delivering value for the user. This essay about data in its various forms is **not** the life-blood but a contributor. Global public goods that define "life-blood" are food, energy, water, sanitation and healthcare (FEWSH). Separately, and in combination, the 80% world need advanced and affordable array of tools and technologies to leapfrog the conventional practices of FEWSH in the 20% world.

In this context energy is the rate limiting entity and in a "tie" with food and water, in terms of human existence and life. The "hand-me-downs" from the 20% world of energy may not be sustainable and perhaps not even good. Perhaps the Sahara Desert may be a source of energy for creating a global "battery" - an idea[261] triggered by an 1877[262] proposal, in a different context (it was, too, subjected to misrepresentation[263] and mockery[264]). Whether this is a "good" idea or not is *not exclusively* a matter of technological feasibility of implementation or transaction cost of service delivery. The question is, if it is *good for people*. Global public goods are a matter of context for the community as well as the continent. Exploring the cleavage between entrepreneurial engineering innovation and complexities of social egalitarianism requires willingness to recognize, and adapt, among many *different conceptions of a sense of the future*. Not the future deemed appropriate by the 20% world experts. The future is non-linear.
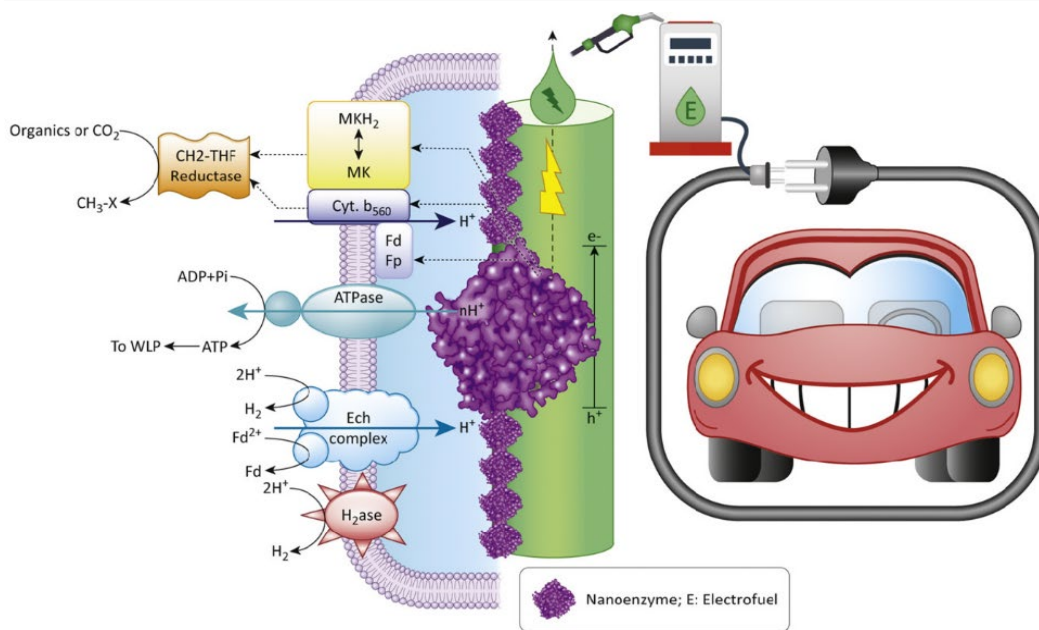

Figure 14 – The trinity of imagination, invention and innovation is central for the 80% world to leapfrog the dead weight of old technology and conventional wisdom from the 20% world. *Nanoenzyme-microbe interaction for clean and **affordable** bio-electrofuel production (Fig 1 from Singh et al [ref 271]).*

It may be a non-binary future with multiple paths and unequal connectivity between amorphous nexus of networks representing non-linear choices, aspirations and outcomes. "Good" decisions are relative to **that mix** which defies definition yet works as a catalyst for economic rejuvenation. Even this **type** of "good" will (must) change with time and culture because no one version of good can fit all the world[265]. A binary *outcome*, with exceptions, must not be confused with binary *decision making* because a plethora of non-binary factors can influence the outcome, which may appear as a simple output.

An oversimplified and cherubic example of the latter may resonate with residents of the Boston area. The choice between Mike's[266] and Modern[267], famed confectioners located almost opposite each other on Hanover Street in Boston's North End, is far from binary. The filling in the cannoli, taste of java and the length of the queue, are factored in the decision-making process, which generally presents itself cloaked in a binary-esque outcome.

Thus, the creativity and imagination necessary from science, engineering, technology and medicine (STEM) to address FEWSH must extract solutions sustainable under economic constraints. They will vary in their need to connect a few or many 'dots' to inform the solution delivery system. It is true "porosity" may contribute to solutions in the 80% world, and perhaps, less is more, but it will be remiss to leave the reader with the impression that invention/innovation may have to take a second place in the 80% world. In some cases, the latter rings true. In most instances we must seek *out-of-the-world* or counterintuitive ideas[268] and blend it with incisive insight. Far reaching *convergence* of bio, nano, info and eco[269] is not an alternative but an imperative to stitch practical solutions, which can satisfy, survive, and surpass, the criteria dictated by the economics of technology and technology policy[270], which may be necessary to transform grand visions[271] into everyday reality (and uncover new[272] tools, in the process).

But "grand visions" are often manufactured[273] as incremental mediocrity. Patents for using quantum[274] computing[275] are as absurd as the use of "cognitive" and the accompanying belligerence[276] in marketing. Those who throw around the term "cognition" may not have consulted a credible expert[277] or explored its meaning/definition {*cognition [n] mental action or process of acquiring knowledge and understanding through thought, experience, and the senses*}. Neither neurology nor modern computational neuroscience comprehends the *combined* electrochemical, cellular and molecular nature of what it may mean to "acquire knowledge" by animals or humans. Any model, equation, algorithm or hand-waving is simply false if it claims anything more than a vague impression of what "*acquire knowledge*" may mean. The other key words in the definition (*thought, experience*) are at depths we do not even dare to know how to measure. The mindless drivel that we can decipher "processes" because of functional nuclear magnetic resonance imaging (fMRI "activation" maps of real time blood flow) are constrained by spatio-temporal resolution and limited ability of fMRI[278] to reliably detect functional activation. At the current state of instrumentation, resolution is inversely proportional to the ability to detect functional activation. Optimizing both is essential before fMRI data may be considered precise.

The infinite absurdity cryptic in the claims about cognition, learning, experience and thought is neither coloured by cognitive dissonance nor a figment of our uninformed imagination. Table 4[279] captures the duration of so-called "deep learning" training over "days" on a TPU[280] (tensor processing unit) scale with vast amounts of data (GB) only to generate undifferentiated[281] rubbish. By comparing row number 1 vs 10 (bottom), the scores of the relevant match (#1, 0.892) between learning (saved query) vs challenge (new query) is unimpressively different (#10, 0.765). In other words, according to Google BERT, after several days of "deep learning" *"Blah blah blah blah"* when challenged with the query *"Does this integrate with gmail?"* generates 0.765, suggesting *76.5% similarity* between the two.

	BERT	RoBERTa	DistilBERT	XLNet
Size (millions)	**Base**: 110 **Large**: 340	**Base**: 110 **Large**: 340	**Base**: 66	**Base**: ~110 **Large**: ~340
Training Time	**Base**: 8 x V100 x 12 days* **Large**: 64 TPU Chips x 4 days (or 280 x V100 x 1 days*)	**Large**: 1024 x V100 x 1 day; 4-5 times more than BERT.	**Base**: 8 x V100 x 3.5 days; 4 times less than BERT.	**Large**: 512 TPU Chips x 2.5 days; 5 times more than BERT.
Performance	Outperforms state-of-the-art in Oct 2018	2-20% improvement over BERT	3% degradation from BERT	2-15% improvement over BERT
Data	16 GB BERT data (Books Corpus + Wikipedia). 3.3 Billion words.	160 GB (16 GB BERT data + 144 GB additional)	16 GB BERT data. 3.3 Billion words.	**Base**: 16 GB BERT data **Large**: 113 GB (16 GB BERT data + 97 GB additional). 33 Billion words.
Method	BERT (Bidirectional Transformer with MLM and NSP)	BERT without NSP**	BERT Distillation	Bidirectional Transformer with Permutation based modeling

	Saved Query	New Query	BERT Score	USE Score	ELMO Score	XLNet Score
1	How much will this cost?	Is this expensive?	0.892	0.803	0.742	0.720
2	Where is your data stored?	How secure is your product	0.890	0.625	0.765	0.705
3	What temperature is it today?	What time is the game on?	0.880	0.677	0.746	0.730
4	How do I change my password	I can't find the settings page	0.868	0.671	0.717	0.706
5	Can I sign up for a free trial	Do I need a credit card to get started?	0.868	0.736	0.753	0.759
6	Where can I view my settings	Does this integrate with gmail?	0.866	0.620	0.717	0.691
7	I really do not like this product	I really like this product	0.865	0.747	0.864	0.870
8	What is the Capital of Ireland?	What time is the film in the cinema	0.865	0.578	0.663	0.778
9	Hello, is there anyone there?	What time is the game on?	0.832	0.594	0.680	0.723
10	Blah blah blah blah	Does this integrate with gmail?	0.765	0.519	0.585	0.688

Table 4 – Google BERT thinks *"blah blah blah"* is **76.5% similar** to *"Does this integrate with gmail?"* The laughable outcome is not at all surprising despite the hordes of brilliant scientists working to create tools (RNet, XLNet, ELMo, BERT, ALBERT, DILBERT, ROBERTA) over the past 20 years because it is gnarly to capture *semantics* of language which has evolved over the past 200,000 years.

The much-ballyhooed *context-awareness* of ELMo[282], BERT[283], and its cousin ALBERT[284], due to permutations and combinations of including masked language model (MLM) and next sentence prediction (NSP), is left in the dust, based on examples in Table 4. Except for those feigning ignorance and *blind-to-reason-fanatics,* observers may not be incorrect in thinking that throwing data (please see "Data" row in the upper part of Table 4) or using generic high volume of data for training (BERT claims to use "Wikipedia" in column 1 row 4) is ineffective. To be effective, these tools (ELMo, BERT, ALBERT, DILBERT, ROBERTA, etc) must use training data *relevant to the context* of target (search). Context-awareness in the absence of data curation is as fake as claiming that a marble bowl is made of gold. Nevertheless, these advances in search techniques are immense strides[285] but the tools still aren't "intelligent" but dumb as doorknobs. The doorknob does not turn unless one turns it or actuates it, manually or mechanically. Wikipedia as an experimental **control** is a plausible idea. Using curated data for training (ANN) may improve accuracy of search and better guide *informed users* to extract notions of connections and relationships with BERT-esque tools as **supplementary aids** (Tshitoyan *et al*[286] provides supporting evidence, Table 5). These tools are of limited value for non-mission critical applications, for example, recommendations (movies, books, restaurants), weather for entertainment (IBM's Weather Channel) and fault tolerant uses (open garage door, on/off sprinklers). Non-essential human-centric uses (congestion routing, temperature control, voice message to email) may qualify if the outcome is almost correct in 80% of the cases. Actual use with humans-in-the-loop (healthcare, emergency response, security) may be scuppered if based on any credible risk versus reward analytics, except for offering *non-binding and non-executable* suggestions or alerts for human decision makers.

Text corpus	Materials	Grammar	All	Corpus size
Wikipedia	2.6	72.8	51.0	2.81B words
Wikipedia elements	2.7	72.1	41.4	1.08B words
Wikipedia materials	2.2	72.8	41.3	781M words
All abstracts	43.3	58.3	51.0	643M words
Relevant abstracts	48.9	54.9	52.0	290M words
Pre-trained model	10.4	47.1	30.8	640k papers

Table 5 (from Tshitoyan *et al* Extended Data Table 4) – Analogy scores (%) for materials science versus 'grammar' from different sources. Training using Wikipedia for – metals – is grammatically rich (>72%) but content poor even when using select Wikipedia for materials (2.2% analogy). The smallest corpus (290M words) used for training using continuous bag of words (CBOW) offers the best performance (48.9%) on materials-related analogies when **curated** for "relevant" abstracts. The best performance for grammar may turn out to be profitable by enabling ELMo, BERT, ALBERT, ROBERTA and DILBERT (DistilBERT) to be the voices of artificial trainers for the standardized twaddle marketed with impunity and known as *Test of English as a Foreign Language* (TOEFL). The contextual enrichment in this table is similar to the example of enrichment shown in Figure 16, suggesting the need for relevance / curation.

The paramount significance of curated contextual data in training any model (including ANNs, artificial neural networks) cannot be overemphasized. Individuals and institutions in possession of less than lofty ideals may revert to trickery in an attempt to sow doubt or discombobulate or disqualify the type of outcomes, for example, presented in Table 5. It is the age-old deception due to over-fitting[287] which can be also applied to ANN during training and the "fit" may be driven to precision using tools such as recursive feature addition[288] (RFA). For readers seeking a simpler analogy may wish to re-visit what we discussed as the "force-fitting" of RDF to healthcare applications (reference 170). The erudition necessary to train ANN with curated data is not easily gleaned from a cursory review. Extensive perusal of scholastic research[289] begins to reveal a few of the minutiae with respect to the nature of the domain specific data and the context of data curation (see sections 3 and 4 in Nandy *et al*, reference 289) that forms the bulk of the *preparatory* work based on rigor and strength of broad spectrum[290] knowledge. In an earlier section we referred to *domain specific* models in a "macro" sense (pages 9-10) whereas the domain specificity of this example (Nandy *et al*) is at the molecular (atomic and/or sub-atomic) scale.
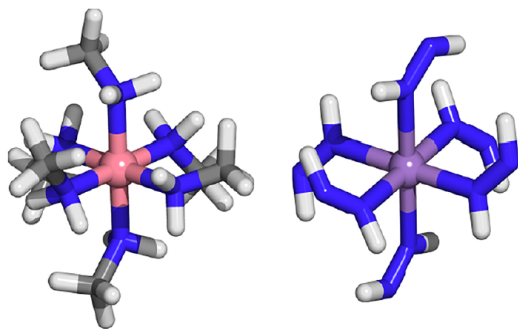
	$[Co(NH_2CH_3)_6]^{3+}$	$[Mn(HNNH)_6]^{3+}$
DFT	-20.00 eV	-18.64 eV
ANN	-19.91 eV	-23.55 eV

Table 6 – Even after extensive training using precision data enriched for features using RFA, it is not surprising when gross errors are found in the outcome (analysis). Figure 7 (from page 13981 in Nandy *et al*) is one example how artificial neural networks (ANN) used in machine learning (ML) exercises and analytics generate erroneous results. ΔE_g data (LEFT) shows ANN error (0.09 eV) with respect to DFT (density functional theory[291]) in a singlet [Co(NH2CH3)6]3+ transition metal complex.

(Right) Data shows large ANN error (−4.91 eV) with respect to DFT for a quintet [Mn(HNNH)6]3+ transition metal complex. The quintet [Mn(HNNH)6]3+ complex highest occupied molecular orbital (HOMO) level is underestimated by 4.9 eV, which is almost ***double*** the mean absolute error (MAE). This ANN was specifically trained using ΔE_g data models on a set of 64 octahedral homoleptic complexes (OH64). The discrepancy (ANN error) is significant because frontier molecular orbital energetics provide essential insight into chemical reactivity and dictate optical and electronic properties. Small errors could make an immense difference in terms of chemistry of the transition metal complex. In this illustration, the metals are shown as spheres and coordinating atoms as sticks (C atoms, gray; N atoms, blue; H atoms, white). *If your healthcare diagnosis and treatment was based on such an ANN outcome, would you trust, accept and abide by the direction of the treatment suggested by such results? If this outcome is based on data from your electronic health records (EHR) which is known to be erroneous, would you trust poor data quality to inform a poorly performing ANN engine to design your healthcare?*

The third piece of evidence that also dispels the marketing myths of AI in favor of viewing through the lens of **artificial reasoning** tools (**ART**, referring to ANN, CNN, RNN, DL, RL), is another variety of neural network[292] with credible capabilities. MPNN[293] (message passing neural network) for molecules[294] is a tool[295] to unleash data[296] for human-centric applications in health and medicine. This example centers on uncovering and repurposing a previously known molecule as an antibiotic[297] using a plethora of tools including MPNN and collectively referred to as deep learning (DL). This paper (297) by Stokes *et al* and the two other papers by Tshitoyan *et al* as well as Nandy *et al*, emphasize data curation and learning, without once mentioning the term AI or "artificial intelligence" in any shape or form in the scientific papers. Unfortunately, the marketing and news item[298], as expected, did not shy away from fake sensationalism to bolster the false appeal of AI.

The *learning* that generated the antibiotic (renamed Halicin), is nauseatingly detailed and the *training* (MPNN) was excruciatingly structured, optimized and re-optimized (using hyperparameter[299] optimization). The old idea of ensembling[300] was applied to improve outcomes *in silico* but predictions were *biologically* tested through rigorous experiments. Even after repeated steps to minimize errors, the authors (reference 297) remain cognizant of the pitfalls: *"It is important to emphasize that machine learning is imperfect. Therefore, the success of deep neural network model-guided antibiotic discovery rests heavily on the coupling of these approaches to appropriate experimental designs."* (Stokes *et al*, page 698)
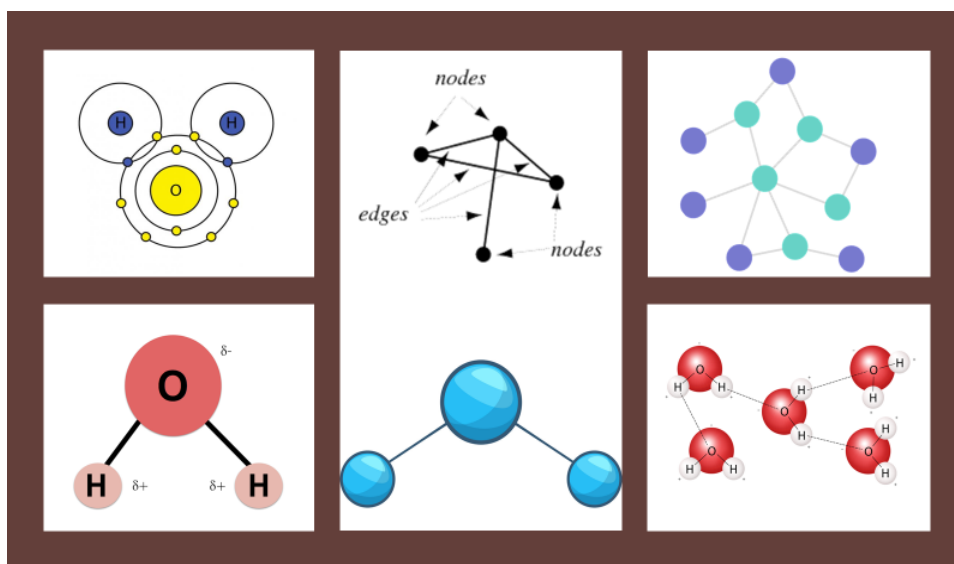
Figure 15 – Training a neural network to recognize molecules relies on the fact that every molecule may be represented as a *graph* (or a collage of connected graphs, eliciting the idea of a *knowledge graph*). The water molecule may be viewed as a graph with oxygen (O) as the node (vertex). Bonds between oxygen and hydrogen (O–H) serves as the "side" or edge. Most molecule (within reason) may be transformed to a molecular graph and is at the heart of MPNN training to recognize different types of molecules. Then, the *trained* neural network, MPNN, is used to search for similar or *dissimilar* molecules in a repository.

A curated set of 2335 molecules were used as the training set for new antibiotic molecules. The 2335 training data set were chosen, including FDA library of 1,760 molecules pre-selected based on their ability to inhibit microbial (*E. coli* BW25113) growth. In other words, molecules with structure and function ***known*** to possess anti-microbial activity. Training MPNN with this data set enables the neural network to *learn* the structures in order to select similar (or dissimilar) structures from a larger library of structures. The expectation is that when a "challenge" library is presented to the MPNN, the degree of similarity or dissimilarity, in terms of the output from the MPNN, can be *tuned* by modifying selection parameters. For example, using prediction scores (PS) to categorize molecules from a larger library (in this case, the ZINC database with ~1.5 billion molecules). By selecting higher PS value (>0.7, >0.8, >0.9), the outcome is "enriched" and a sub-set of molecules (in this case, 107,347,223, reductionism at work) is further subjected to other selection criteria, for example, nearest neighbor analysis (Tanimoto score). Finally, potential molecules (in this case, 23) are biologically screened (microbial assay) to identify the "new" antibiotic candidate(s). One such candidate is Halicin (Stokes *et al*), previously identified as the c-Jun N-terminal kinase inhibitor SU3327 and re-discovered as a broad-spectrum antibiotic, re-named Halicin but still the same molecule as SU3227, albeit repurposed, based on function.

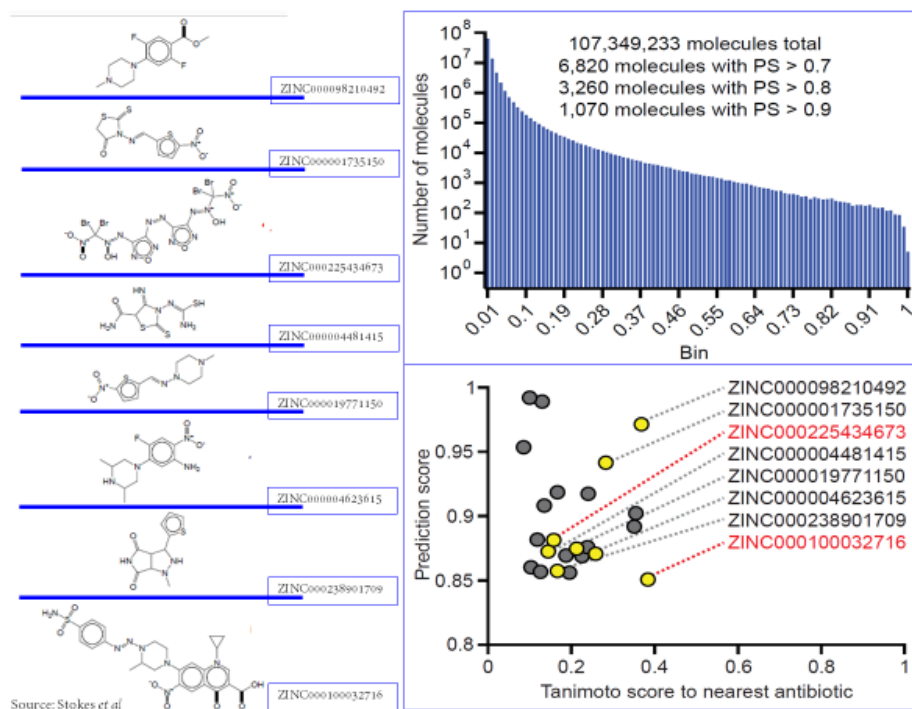Figure 16 – Potential candidates (8 molecules) from ZINC database (structures on the left) were scored using nearest neighbor (NN) analysis (yellow circles, bottom right). NN is based on principles derived a thousand years ago[301] (circa 1030). Data is curated at successive steps by enriching for ***context*** (selecting higher prediction scores, PS, top right) in a manner similar to Table 5 (wikipedia vs relevant abstracts).

In combination, these three examples offers preliminary evidence that artificial reasoning tools (ART) such as ANN, MPNN, DL (deep learning), RL (reinforcement learning), etc. are **excellent** tools. However, ART and related tools are not intelligent, they do not self-operate and the outcome is solely due to the skill and sophistication of the human operators. The steps must be designed with cautious intellectual strategy embracing the **breadth** of diverse knowledge, often dismissed by many institutions. Execution demands **depth** of erudition and incisive foresight to weigh the pros and cons of the criteria used to assess the *quality of curated data* prior to commencing training neural networks with such data.

It is essential to learn the *meaning of context* in order to sufficiently inform the "artificial" part of ART. Models and patterns are like chicken playing tic-tac-toe[302] without context and semantics. Human *knowledge* to equip ART is almost impossible to transfer because we do not have a clue how to abstract continuous knowledge and use discrete processes to **build** it into an artificial system. Hence, ART may not "possess" an internal model of the external world. The immense variability in terms of features and which features may be *relevant* in which environment makes it difficult to model a state by claiming that feature selection will address all relevant and discrete contexts that the item or object may experience. Even if feature engineering was automated to levels of precision continuity that could encapsulate all possible permutations and combinations of the behavior of an entity or object, the model may be inadequate in the hands of different users who may not feature in the feature catalog, without bias. It is not trivial yet not impossible to model behavior and optimize for some features in a retail environment (for example, who may shop at Whole Foods, who may return to Andronico's versus Mollie Stone's).

If reason could inform common sense then one may prefer ART over AI and the observe the value of reasoning in machine learning techniques using neural networks. Neural networks and machine learning tools are amplifying, modifying and regurgitating whatever humans have programmed into the tool. It cannot **learn** beyond the range of data or information provided, until humans decide to change, adjust or add/subtract parameters/attributes which will influence the **learning** and the output from ART. The obstreperous zeal to move away from the misnomer of modern[303] AI[304] and adopt ART as a generic term may be a *back to the future* moment for rule-based[305] expert systems[306] and principles[307] but coupled with new ML tools[308]. Marketing panders to creators of unstructured data but the zettabytes of data anarchy occasionally offers value irrespective of the clamor for general AI or ambient AI or intuitive AI or cognitive AI. Isn't it possible to deliver value using ART?

But, does the acronym matter? Perhaps only to a few (or less). AI is a false trigger for technology transitions[309] but it is cheaper[310] and cheaper[311] to promote. It is profitable for conference organizers, narcissistic speakers, greedy social gurus and other forms of eejits. Irresponsible computation may be draining the energy[312] economy yet the marketing world is oblivious to the grave socio-economic incongruencies in terms of the thermodynamics[313] of computation, which is absent from daily discussions. Perpetrating the myth of intelligence in AI is a moral anathema. ART lacks the cachet and panache, but promotes the rational idea of *learning* tools, which are, and will be, helpful for society.

The state of artificial learning is analogous to receiving a map of the world on a postage stamp and expect the bearer of the map (stamp) to arrive at 77 Massachusetts Avenue, Cambridge, MA, using that postage-stamp-sized map as the only guide. Neurologists shudder[314] at AI, the public are ignorant of the evidence of sham (Tables 4-6 and Figure 16) while marketing accelerates the "show" over substance. Sensationalism amplifies attention and siphon funds away from real world issues, making it harder for elements of FEWSH[315] to move forward, for the 80% world. The task ahead is to be creative, more than expected, and avoid the oxymoronic implementation of *innovation as usual*. Dynamic combinations[316] and cross-pollination of counterintuitive connections may be worth exploring[317] to find many *different* ways to lift billions of boats, not just a few yachts. Future needs égalitarian resistance to our "default" state of society where we allow, acquiesce and accept greed[318] as a price and penalty we pay for progress.
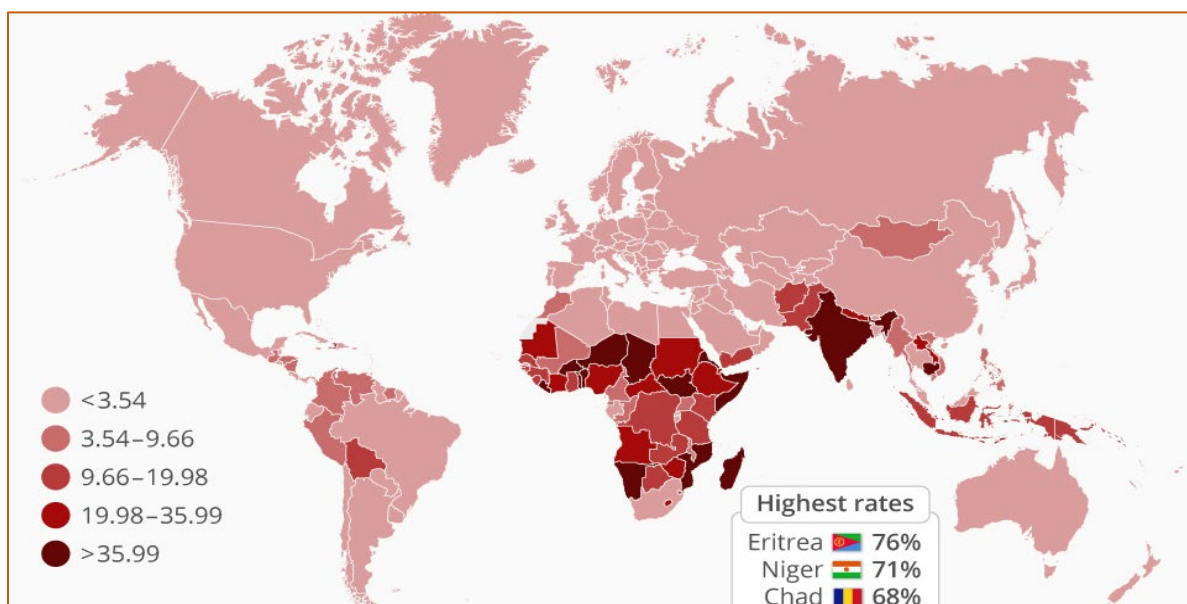
Figure 17 – Is GINI coefficient[319] a non-stationary end-goal for economic redistribution through ethical social entrepreneurial innovation? Nearly a billion[320] people[321] defecate outdoors (percent of population who are forced to defecate outdoors). It appears that the pay-a-penny-per-unit (PAPPU) model could rake in billions if managed sanitation services were developed as a business. If a billion people paid one penny (US) per use per day for their "leased" sanitation service (at home) then the global gross earning for pay-per-use sanitation may be US$3.65 billion annually, an indication of earnings potential and wealth from the business of the poor. The primary assumption is that the individual will choose to pay one penny per day even if their income is only $2 per day (lowest per capita average earnings). This social business model depends on different domains[322] of infrastructure necessary to offer home sanitation as an e-commerce[323] service. The return on investment will be realized gradually because earnings will not be US$3.65 billion in the first year. Is the inclination to invest, and wait, too much to expect from global organizations which could help facilitate delivery of global public goods?

China 🇨🇳	528,263	20.67%
United States 🇺🇸	422,808	16.54%
India 🇮🇳	135,788	5.31%
Germany 🇩🇪	104,396	4.08%
Japan 🇯🇵	98,793	3.87%
United Kingdom 🇬🇧	97,681	3.82%
Russia 🇷🇺	81,579	3.19%
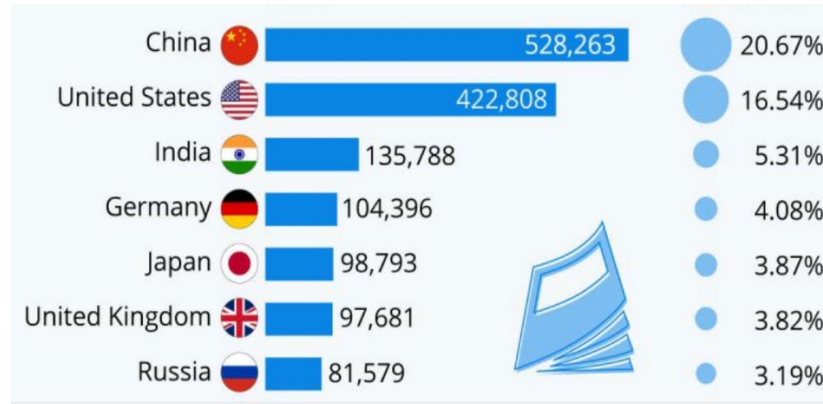
Figure 18 – Publications in peer-reviewed journals (2018). Face-saving feel-good false positives? Propaganda[324] (does **not** exclude scientists[325]) masks facts[326] and analyses, by ignoring the quality of publications, citations and investment in R&D (% GDP). Dubious research[327] output tarnishes the image and publications. Are we pointing fingers at a quarter (25.98% = 20.67% + 5.31%) of the global share?

Photograph 0 – GINI coefficient gone awry? Vast slums, adjacent to high rise residential buildings in a section of Mumbai (photograph by Prashant Waydande[328]), contradicts the notion that India may be an emerging leader in credible scientific research[329] (Figure 18). Are these a few symptoms stemming from grave gender bias, discrimination[330] against females and inequity of women in science and society?

PERSONAL END NOTE

Thoughtful readers have asked about the audience[331] that this essay intends to reach. I am unable to answer. This essay flows from convergence of a broad spectrum of chatter. Certain segments originate from personal observations and others are a mix of intent, content, context and relevance from social network of users as well as decision makers[332]. Most people prefer focus and a *take home message*. The former is less obvious but the latter is less unclear from the cartoon at the bottom of this page. It may be uninteresting for most adults to read about an array of haphazard realities, an extrapolation from my exchanges (or lack thereof) with a few hundred selected professionals in academia and industry. But, what is more painful is the absence of engagement with young minds and students. Isn't it tragic to be young and cynical? Has the person gone from knowing little to exploring even less? Ruminations in this essay may not reach more than a few people. The fabric of this essay was spun with many threads, not all of them are digital. To arrive at the conjectures in this essay consumed part of my life in their acquisition and formation, irrespective of the fact that they be incorrect, flippant, vacant or just rubbish. Thus, one may conclude, perhaps the essay exposes my stupidity[333], perhaps it is irrelevant and serves no real purpose, perhaps *it is a tale*[334]*, told by an idiot, full of sound and fury, signifying nothing.*

Les savants des autres nations, à qui nous avons donné l'exemple, ont cru avec raison qu'ils écriraient encore mieux dans leur langue que dans la nôtre. L'Angleterre nous a donc imités; l'Allemagne, où le latin semblait s'être réfugié, commence insensiblement à en perdre l'usage; je ne doute pas qu'elle ne soit bientôt suivie par les Suédois, les Danois et les Russes. Ainsi, avant la fin du dix-huitième siècle, un philosophe qui voudra s'instruire à fond des découvertes de ses prédécesseurs, sera contraint île charger sa mémoire de sept à huit langues différentes, et. après avoir consume à les apprendre le temps le plus précieux de sa vie, il mourra avant de commencer à s'instruire.[335]	The scholars of other nations, to whom we have provided an example, believed with reason that they would write even better in their language than in ours. England has thus imitated us; Germany, where Latin seems to have taken refuge, begins insensibly to lose the use of it: I do not doubt that it will soon be followed by the Swedes, the Danes and the Russians. Thus, before the end of the 18th century, a philosopher who would like to instruct himself about his predecessor's discoveries will be required to load his memory with 7 to 8 different languages; and after having consumed the most precious time of his life in acquiring them, he will die before having begun to instruct himself.[336]

If it's inaccessible to the poor it's neither radical nor revolutionary.
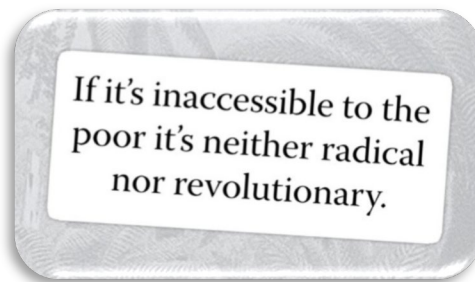
IN CONCLUSION – THE PRINCIPLES AND PRACTICE OF P3 – IS IN NEED OF PEOPLE WHO ARE TRUE LEADERS

Concluding sections ought to be labeled "temporary" because we are perpetually immersed in an ocean of unknown unknowns. This essay and our discussion is not an exception. Nevertheless, it will be remiss to leave the discussion without an attempt to share clues with entrepreneurial innovators who may be interested in social businesses and, in parallel, the pursuit of ethical profitability. Perusal of the PDF documents "Practice of P3" will reveal a broader context and perhaps a few ideas related to food and water (01FW), healthcare (02H) and energy (03E) - https://dspace.mit.edu/handle/1721.1/123984

As the name implies, social businesses are "***businesses***" and not academic experiments or lab projects. The convergence of concepts, tools and technologies necessary to deliver the business of social business is not trivial and the most important element is that of ***leadership***. Convergence implies that collaboration between domains rich in ideas, integration and implementation are, more or less, essential for end-to-end service delivery. No <u>***one***</u> business can synthesize the entire tapestry required to deliver a certain quality of service. Each business (and academic partners) must contribute to the ecosystem and relinquish certain elements of control the business executives assume. The service delivery ecosystem is comprised of many types of businesses and some may be inclined to conclude that they are the "pillars" in this new process. Too often one forgets that pillars are useless unless we can build bridges over pillars. This *forgetfulness* is buoyed by hubris, mostly at the CxO level, where individuals designated "chief officers" are ill-equipped to navigate the paradigm shift to *service* delivery. In particular, ***people*** who are CTO or CSO or CIO or CDO may think that they possess all the tools and resources to provide what is best. They may know what is best for their domain or company but it may not be the same as what is best for the end user in the ecosystem and the service supply chain partners in the delivery network.

The CxO roles must evolve to keep pace with emerging spectrum of models where products are pariahs and service is the new slogan. Management roles ought to change (CEO, COO, CFO). Urgent changes are overdue for CTO, CSO, CIO, CDO roles (excludes people who maintain IT plumbing). In the world of the "outcome economy" selling a trainer or a sneaker is no longer the end point. The trainer or sneaker is now a "service" where the consumer pays for duration of the service (life cycle) provided by the trainer or sneaker. It may seem tad obsequious but the service economy is here to stay, especially for the 80% world. The *nano-fee pay-a-penny-per-use* (PAPPU) model may unlock markets of billions who may have non-zero amounts of disposable income. Wealth of the poor may fuel ethical profitability for social business by tapping into markets of the next billion users. Current shades of 'western' business ethos, roles and models are unlikely to serve as pillars for the bridge to the future of service economies. The insurmountable resistance to reducing *western* business barriers is comparable to the resistance from the American Medical Association to universal healthcare. *However, exceptions may be found in cases where the CEO or CTO is/are the primary inventors (creating start-ups). The question for them is whether can they successfully and seamlessly integrate their contribution with the service supply chain.*

In general, one of many errors in the current CxO model is the focus on specialization, which is highly relevant to CTO, CSO, CIO, CDO, because the company values the skills the person contributes. While that skill is certainly useful and valuable to a segment of the business, it may be less applicable when that skill is only a part of the solution. The outcome economy does not care if the CTO is skilled in millimeter wavelets or the CDO is an expert in backpropagation algorithms. The ***delivery of value*** at the end of the tunnel is not about a skill, but a synergistic integration of a *variety of skills*, most of which may be missing in any one person and in any one company. The idea of a generalist gently rises to the surface.

The future is about fusion. The rate limiting nodes are due to the nature of the CxO roles who are, traditionally, pefer to drown in their own knowledge, even if the facts are out of sync. The new world CxO roles may emphasize coordination and *cross-pollination*. The CxO must be a dyamic **_node_** who can connect the company to access knowledge as well as resources from a wide cross-section of academic, industry and government sources, depending on the case. This is important, immediately, to CTO, CSO, CIO, CDO roles because of the complex and inextricably linked dependencies, at the core of service delivery solutions. This change is not only for social businesses but for every business, especially small and medium businesses, where resources are limited yet the solutions demand increasing degrees of sophistication. This change is neither a silver bullet nor is it fitting for all intents and purposes.

Part of the disconnect, in Fortune 500 type companies, may be related to the salaries of CxO's. For example, if a half-million-dollar CTO seeks outside help for a project, the CEO may wonder why the company should yet again pay for external experts when it has already paid half-million for the "know-it-all" CTO. The depth of the in-house talent may be inadequate. At the end, a half-baked outcome is served to the customer, who is brain-washed to believe that the innovation delivered exceed all others.

For human resources, hiring a CTO,CSO, CIO, CDO as a "learn-it-all" **_node_** for knowledge connectivity may be heretic. The constant need for creating liaison may be unacceptable. Most CxO individuals may not have the correct network or may not know who knows what, where. Composing and re-composing solution teams through agile combinations of internal and external relationships are rarely in practice. Yet, it is the latter that is poised to deliver exponential value, if explored. Few **_leaders_** have the incisive foresight to lead beyond their comfort zone. Even fewer possess the courage to take the road less travelled. The system appears to reward the *risk-averse* managers (albeit, with exceptions).

Businesses will benefit from people/nodes of connectivity who can act as **_coordinators_**. Social businesses who aim to be profitable will require immense connectivity. Coordination must be the daily mantra. Perusal of "Practice of P3" will reveal that trans-disciplinarity is the norm. The definition of CxO roles must be deconstructed and reconstructed as an amalgam of parts, if we are serious about service delivery of global public goods, for example, food, water (01FW), healthcare (02H) and energy (03E) - the crucial elements of our haphazard reality (https://dspace.mit.edu/handle/1721.1/123984).

REFERENCES

[1] https://en.wikipedia.org/wiki/Pareto_principle

[2] https://www.postscapes.com/iot-history/

[3] Sanjay Sarma, David Brock and Kevin Ashton (1999) *"The Networked Physical World - Proposals for Engineering the Next Generation of Computing, Commerce, and Automatic-Identification,"* MIT Auto-ID Center White Paper. MIT-AUTOID-WH001, 1999. https://autoid.mit.edu/publications-0 https://pdfs.semanticscholar.org/88b4/a255082d91b3c88261976c85a24f2f92c5c3.pdf

[4] Sarma, S., Brock, D. and Engels, D. (2001) "Radio Frequency Identification and the Electronic Product Code," *IEEE Micro*, vol. 21, no. 6, Dec. 2001, pp. 50–54. doi:10.1109/40.977758.

[5] Sarma, Sanjay (2001) "Towards the 5 cents tag," Auto-ID Center. MIT-AUTOID-WH-006, 2001

[6] https://www.rfc-editor.org/rfc/pdfrfc/rfc791.txt.pdf

[7] Deering, S. and Hinden, R (2017) Internet Protocol, Version 6 (IPv6) Specification. Internet Engineering Task Force (IETF) https://www.rfc-editor.org/rfc/pdfrfc/rfc8200.txt.pdf

[8] https://www.bizjournals.com/boston/blog/mass-high-tech/2012/11/mit-aims-to-harness-cloud-data-on-consumer.html

[9] Max Mühlhäuser and Iryna Gurevych (2008) Introduction to Ubiquitous Computing. IGI Global. https://pdfs.semanticscholar.org/ab0e/b44c7c81a1af3fc2d23fa03f8f04f9e4ca2d.pdf

[10] https://www.bigdataframework.org/short-history-of-big-data/

[11] Melina Kourantidou (2019) *Artificial intelligence makes fishing more sustainable by tracking illegal activity.* https://theconversation.com/artificial-intelligence-makes-fishing-more-sustainable-by-tracking-illegal-activity-115883

[12] https://www.postscapes.com/iot-consulting-research-companies/

[13] Which market view? The imperfect categories are consumer and industrial IoT. Definitions proposed by domestic and foreign governments (US DHS, US FTC, EU ENISA) are (not surprisingly) neither systematic nor in sync. According to DHS, IoT is defined as "connection of systems and devices (with primarily physical purposes e.g. sensing, heating/cooling, lighting, motor actuation, transportation) to information networks (internet) via interoperable protocols, often built into embedded systems." (page 2 in "Strategic Principles for Securing the Internet of Things (IoT)", Version 1.0, Nov 15, 2016) https://www.dhs.gov/sites/default/files/publications/Strategic_Principles_for_Securing_the_Internet_of_Things-2016-1115-FINAL_v2-dg11.pdf).
European Union Agency for Network and Information Security (ENISA) defines IoT as "a cyber-physical ecosystem of interconnected sensors and actuators, which enable intelligent decision making." Internet viewed as a subset not the whole set. (Page 18 in "Baseline Security Recommendations for IoT" https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot/at_download/fullReport). One version of IoT is in H.R.1668 "Internet of Things Cybersecurity Improvement Act of 2019" (https://www.congress.gov/116/bills/hr1668/BILLS-116hr1668ih.pdf).

H.R. 4792 "U.S. Cyber Shield Act of 2019" mentions "internet-connected products" and stipulates that the term 'covered product' means a consumer-facing physical object that can (a) connect to the internet or other network; and (b) (i) collect, send, or receive data; or (ii) control the actions of a physical object or system (https://www.congress.gov/116/bills/hr4792/BILLS-116hr4792ih.pdf).

[14] Cisco CEO at CES 2014: *Internet of Things is a $19 trillion opportunity*.

https://www.washingtonpost.com/business/on-it/cisco-ceo-at-ces-2014-internet-of-things-is-a-19-trillion-opportunity/2014/01/08/8d456fba-789b-11e3-8963-b4b654bcc9b2_story.html

[15] https://www.the-scientist.com/uncategorized/nature-rejects-krebss-paper-1937-43452

[16] www.zdnet.com/article/internet-of-things-8-9-trillion-market-in-2020-212-billion-connected-things/

[17] https://machinaresearch.com/static/media/uploads/machina_research_press_release_-_m2m_global_forecast_&_analysis_2012-22_dec13.pdf

[18] https://www.npr.org/sections/thesalt/2017/07/20/527945413/khichuri-an-ancient-indian-comfort-dish-with-a-global-influence

[19] Mendel, Gregor. 1866. Versuche über Plflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen, 3–47.

http://www.esp.org/foundations/genetics/classical/gm-65.pdf

[20] F. Jacob and J. Monod, "Genetic regulatory mechanisms in the synthesis of proteins," *Journal of Molecular Biology*, vol. 3, no. 3, pp. 318–356, 1961. https://doi.org/10.1016/S0022-2836(61)80072-7

[21] Dulbecco, Renato. "The Induction of Cancer by Viruses." *Scientific American*, vol. 216, no. 4, Apr. 1967, pp. 28–37 doi:10.1038/scientificamerican0467-28

http://calteches.library.caltech.edu/230/1/cancer.pdf

[22] Novikov, D. A. "Systems Theory and Systems Analysis. Systems Engineering." *Cybernetics*, by D.A Novikov, vol. 47, Springer International Publishing, 2016, pp. 39–44. doi:10.1007/978-3-319-27397-6_4 www.researchgate.net/publication/300131568_Systems_Theory_and_Systems_Analysis_Systems_Engineering

[23] http://web.mit.edu/esd.83/www/notebook/Cybernetics.PDF

[24] Wiener, Norbert. *Cybernetics, or Control and Communication in the Animal and the Machine (2nd Ed.)*. MIT Press, 1961. doi:10.1037/13140-000.

https://uberty.org/wp-content/uploads/2015/07/Norbert_Wiener_Cybernetics.pdf

[25] https://www.sissa.it/fa/workshop_old/DCS2003/reading_mat/zuazuaDivSEMA.pdf

[26] Williams, L. Pearce. "André-Marie Ampère." *Scientific American*, vol. 260, no. 1, 1989, pp. 90–97.

https://www.jstor.org/stable/24987112

[27] https://www.americanscientist.org/article/alice-and-bob-in-cipherspace

[28] Rivest, R. L., et al "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems." *Communications of the ACM*, vol. 21, no. 2, Feb. 1978, pp. 120–26 doi:10.1145/359340.359342. https://dl.acm.org/citation.cfm?id=359342

[29] Nash, J. F. (1950) "Equilibrium Points in N-Person Games." *Proc Natl Academy of Sci (USA)*, vol. 36, no. 1, Jan. 1950, pp. 48–49. doi:10.1073/pnas.36.1.48 ▪ www.pnas.org/content/pnas/36/1/48.full.pdf

[30] Axelrod, Robert M. *The Evolution of Cooperation*. Rev. ed, Basic Books, 2006. http://www.eleutera.org/wp-content/uploads/2015/07/The-Evolution-of-Cooperation.pdf

[31] Kreps, David M., et al. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory*, vol. 27, no. 2, Aug. 1982, pp. 245–52. https://doi.org/10.1016/0022-0531(82)90029-1

[32] https://en.wikipedia.org/wiki/Graph_database#/media/File:GraphDatabase_PropertyGraph.png

[33] Little, John D. C., et al. "An Algorithm for the Traveling Salesman Problem." *Operations Research*, vol. 11, no. 6, Dec. 1963, pp. 972–89. doi:10.1287/opre.11.6.972. https://dspace.mit.edu/bitstream/handle/1721.1/46828/algorithmfortrav00litt.pdf

[34] Schrader, Charles R. (1997) United States Army Logistics, 1775-1992: An Anthology, Volume 1 https://history.army.mil/html/books/068/68-1/cmhPub_68-1.pdf

[35] Datta, Shoumen, et al. "Adaptive Value Networks." *Evolution of Supply Chain Management: Symbiosis of Adaptive Value Networks and ICT*, edited by Yoon S. Chang et al., Springer US, 2004, pp. 3–67. doi:10.1007/0-306-48696-2_1. https://link.springer.com/chapter/10.1007/0-306-48696-2_1

[36] www.cs.cmu.edu/afs/cs/academic/class/15251/Site/current/Materials/Lectures/Lecture13/lecture13.pdf

[37] https://kluge.in-chemnitz.de/documents/fractal/node2.html

[38] https://trove.nla.gov.au/work/21161891?q&versionId=25229969

[39] Sterling, Mary Jane (2015) Mathematics and Art. https://www.bradley.edu/dotAsset/d62c7fce-87ed-4b61-9ce7-23a78b70144d.pdf

[40] Carmody, Kevin. "Circular and Hyperbolic Quaternions, Octonions, and Sedenions." *Applied Mathematics and Computation*, vol. 28, no. 1, Oct. 1988, pp. 47–72. doi:10.1016/0096-3003(88)90133-6. https://doi.org/10.1016/0096-3003(88)90133-6

[41] Pennisi, Elizabeth, et al. "The Momentous Transition to Multicellular Life May Not Have Been so Hard after All." *Science*, 28 June 2018. https://www.sciencemag.org/news/2018/06/momentous-transition-multicellular-life-may-not-have-been-so-hard-after-all

[42] "How Many Bacteria Live on Earth?" *Sciencing*. https://sciencing.com/how-many-bacteria-live-earth-4674401.html

[43] Prusiner, S. B. "Novel Proteinaceous Infectious Particles Cause Scrapie." *Science*, vol. 216, no. 4542, Apr. 1982, pp. 136–44. doi:10.1126/science.6801762. https://science.sciencemag.org/content/216/4542/136.long

[44] Pattison, I. H., and K. M. Jones. "The Possible Nature of the Transmissible Agent of Scrapie." *Veterinary Record*, vol. 80, no. 1, Jan. 1967, pp. 2–9. doi:10.1136/vr.80.1.2. https://veterinaryrecord.bmj.com/content/80/1/2

[45] "1.4 Universality." *New England Complex Systems Institute*. https://necsi.edu/14-universality

[46] Denton, Peter B., et al. "Eigenvectors from Eigenvalues." *ArXiv:1908.03795* (August 2019) https://arxiv.org/pdf/1908.03795.pdf

[47] Tao, Terence, and Van Vu. "Random Matrices: Universality of Local Eigenvalue Statistics." *ArXiv:0906.0510* (June 2010) https://arxiv.org/pdf/0906.0510.pdf

[48] Wu, Leting, et al. "A Spectral Approach to Detecting Subtle Anomalies in Graphs." *Journal of Intelligent Information Systems*, vol. 41, no. 2, Oct. 2013, pp. 313–37. doi:10.1007/s10844-013-0246-7 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.499.3228&rep=rep1&type=pdf

[49] Denton, Peter B., et al. "Eigenvalues: The Rosetta Stone for Neutrino Oscillations in Matter." *ArXiv:1907.02534* (July 2019) https://arxiv.org/pdf/1907.02534.pdf

[50] Committee on Information Technology, Automation, and the U.S. Workforce, et al. *Information Technology and the U.S. Workforce: Where Are We and Where Do We Go from Here?* National Academies Press, 2017. doi:10.17226/24649. National Academies of Sciences, Engineering, and Medicine 2017. https://doi.org/10.17226/24649 and https://www.nap.edu/download/24649

[51] Wassén, Olivia. "Big Data Facts - How Much Data Is out There?" *NodeGraph*, 2 Sept. 2019. https://www.nodegraph.se/big-data-facts/

[52] https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

[53] David, Paul A. "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox." *The American Economic Review* 80, no. 2 (1990): 355-61. www.jstor.org/stable/2006600. (https://pdfs.semanticscholar.org/dff7/9b2f28cbb79da91becaab803667f30394233.pdf?_ga=2.215911511.1520557695.1570682035-1238830782.1562127126)

[54] Syverson, C. (2013) Will History Repeat Itself? Comments on "Is the Information Technology Revolution Over?" Intl Productivity Monitor **25** 37-40 www.csls.ca/ipm/25/IPM-25-Syverson.pdf

[55] Syverson, Chad. *Challenges to Mismeasurement Explanations for the U.S. Productivity Slowdown.* Working Paper, 21974, National Bureau of Economic Research, Feb. 2016. *National Bureau of Economic Research*, doi:10.3386/w21974. https://www.nber.org/papers/w21974.pdf

[56] Gordon, R (2016) *The Rise and Fall of American Growth: U.S. Standard of Living since the Civil War* https://press.princeton.edu/books/hardcover/9780691147727/the-rise-and-fall-of-american-growth

[57] Solow, Robert M. "A Contribution to the Theory of Economic Growth." *The Quarterly Journal of Economics*, vol. 70, no. 1, Feb. 1956, p. 65. doi:10.2307/1884513

[58] *"Information Technology and the U.S. Workforce: Where Are We and Where Do We Go from Here?"* doi:10.17226/24649 https://www.nap.edu/read/24649/chapter/5#55

[59] Bivens, Josh and Mishel, Lawrence (2015) Understanding the Historic Divergence Between Productivity and a Typical Worker's Pay: Why It Matters and Why It's Real. Economic Policy Institute. https://www.epi.org/files/2015/understanding-productivity-pay-divergence-final.pdf

[60] "Why Hasn't Technology Sped up Productivity?" *Chicago Booth Review*. https://review.chicagobooth.edu/economics/2018/article/why-hasn-t-technology-sped-productivity

[61] OECD. *OECD Compendium of Productivity Indicators 2019*. OECD, 2019. doi:10.1787/b2774f97-en

[62] Huff, Darrell (1954) *How to Lie with Statistics*. Norton, 1954. http://faculty.neu.edu.cn/cc/zhangyf/papers/How-to-Lie-with-Statistics.pdf

[63] Blum, Avrim, et al. *Foundations of Data Science*. First edition, Cambridge University Press, 2020. https://www.cs.cornell.edu/jeh/book.pdf

[64] David, Javier. "Study: Nearly 70 Percent of Tech Spending Is Wasted." *Vox*, 31 Oct. 2015, https://www.vox.com/2015/10/31/11620222/study-nearly-70-percent-of-tech-spending-is-wasted

[65] Zobell, Steven. "Why Digital Transformations Fail: Closing The $900 Billion Hole In Enterprise Strategy." *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2018/03/13/why-digital-transformations-fail-closing-the-900-billion-hole-in-enterprise-strategy/

[66] Giller, Graham L. "A Generalized Error Distribution." *SSRN Electronic Journal*, 2005. doi:10.2139/ssrn.2265027.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.542.879&rep=rep1&type=pdf

[67] McDonald, James B., and Yexiao J. Xu. "A Generalization of the Beta Distribution with Applications." *Journal of Econometrics*, vol. 66, no. 1–2, Mar. 1995, pp. 133–52. doi:10.1016/0304-4076(94)01612-4. https://www.sciencedirect.com/science/article/abs/pii/0304407694016124

[68] Granger, Clive W.J. (1983) Co-integrated variables and error-correcting models. UCSD Discussion Paper 83-13.

[69] Engle, Robert F. "Interpreting Spectral Analyses in Terms of Time-Domain Models." *Annals of Economic & Social Measurement.* Vol 5, no 1, Jan. 1976, pp. 89–109. www.nber.org/chapters/c10429.pdf

[70] Engle, Robert F. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, vol. 50, no. 4, July 1982, p. 987. doi:10.2307/1912773. http://www.econ.uiuc.edu/~econ508/Papers/engle82.pdf

[71] Granger, Clive W. J., et al. *Essays in Econometrics: Collected Papers of Clive W.J. Granger*. Cambridge University Press, 2001. https://dl.acm.org/citation.cfm?id=781849

[72] Granger, Clive W. J. *Time Series Analysis, Cointegration, and Applications*. 2004. https://escholarship.org/uc/item/2nb9f668

[73] Engle, Robert. "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics." *Journal of Economic Perspectives*, vol. 15, no. 4, Nov. 2001, pp. 157–68. doi:10.1257/jep.15.4.157. http://www.cmat.edu.uy/~mordecki/hk/engle.pdf

[74] https://www.nobelprize.org/prizes/economic-sciences/2003/summary/

[75] Baldassare Fronte, Greta Galliano and Carlo Bibbiani (2016) From freshwater to marine aquaponic: new opportunities for marine fish species production. Conference VIVUS, 20th and 21st April 2016, Biotechnical Centre Naklo, Strahinj 99, Naklo, Slovenia. https://pdfs.semanticscholar.org/f8b2/fad3fe3f6f92a7b1c8276ad57876d8fd0a70.pdf

[76] Matthias, Dipika M., et al. "Freezing Temperatures in the Vaccine Cold Chain: A Systematic Literature Review." *Vaccine*, vol. 25, no. 20, May 2007, pp. 3980–86. doi:10.1016/j.vaccine.2007.02.052. https://www.nist.gov/sites/default/files/documents/2017/05/09/FreezingReviewArticle-Vaccine.pdf

[77] D. Ruppert (2011) *Statistics and Data Analysis for Financial Engineering*. Pages 477-504. Springer Texts in Statistics. DOI 10.1007/978-1-4419-7787-8_18 https://faculty.washington.edu/ezivot/econ589/ch18-garch.pdf

[78] Duan, Jin-Chuan (2001) GARCH Model and Its Application.

http://jupiter.math.nctu.edu.tw/~weng/seminar/GarchApplication.pdf

[79] Diebold, F.X. (2019) Econometric Data Science: A Predictive Modeling Approach

http://www.ssc.upenn.edu/~fdiebold/Textbooks.html

https://www.sas.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf

[80] Datta, Shoumen and Granger Clive (2006) Potential to Improve Forecasting Accuracy: Advances in Supply Chain Management. https://dspace.mit.edu/handle/1721.1/41905

[81] Datta, S., et al. "Management of Supply Chain: An Alternative Modelling Technique for Forecasting." *Journal of the Operational Research Society*, vol. 58, no. 11, Nov. 2007, pp. 1459–69. doi:10.1057/palgrave.jors.2602419. https://dspace.mit.edu/handle/1721.1/41906

[82] Narayanan, Arvind (2019) *How to recognize AI snake oil*

https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf

[83] Li, Lingli, et al. "A Survey of Uncertain Data Management." *Frontiers of Computer Science*, vol. 14, no. 1, Feb. 2020, pp. 162–90. doi:10.1007/s11704-017-7063-z.

https://link.springer.com/article/10.1007/s11704-017-7063-z

[84] Agrawal, R., Imielinski, T., Swami, A. (1993) "Mining Association Rules between Sets of Items in Large Databases." *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD '93*, ACM Press, 1993, pp. 207–16. doi:10.1145/170035.170072.

[85] Coenen, Frans. "Data Mining: Past, Present and Future." The Knowledge Engineering Review, vol. 26, no. 1, Feb. 2011, pp. 25–29. doi:10.1017/S0269888910000378.

https://www.researchgate.net/publication/220254364_Data_mining_Past_present_and_future

[86] Kahneman, Daniel. *Thinking, Fast and Slow*. Penguin Books, 2012.

http://sysengr.engr.arizona.edu/OLLI/lousyDecisionMaking/KahnemanThinkingFast&Slow.pdf

[87] Akerlof, George, and Janet Yellen. "Rational Models of Irrational Behavior." *American Economic Review*, vol. 77, no. 2, 1987, pp. 137–42. https://notendur.hi.is/ajonsson/kennsla2003/Akerlof_Yellen.pdf

[88] https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care

[89] *Minds and Machines Postponed / Canceled - Post Regarding GE Digital Layoffs.*

https://www.thelayoff.com/t/VtbKYAf

[90] http://bit.ly/MRC-BERT-HAN

[91] Rumelhart, David E., et al. "Learning Representations by Back-Propagating Errors." *Nature*, vol. 323, no. 6088, Oct. 1986, pp. 533–36. doi:10.1038/323533a0.

https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf

[92] McCarthy, John, Minsky, Marvin L., Rochester, Nathaniel and Shannon, Claude E. (1955) A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (31 August 1955) http://www-formal.stanford.edu/jmc/history/dartmouth.pdf

[93] Datta, Shoumen and Granger Clive (2006) Potential to Improve Forecasting Accuracy: Advances in Supply Chain Management. https://dspace.mit.edu/handle/1721.1/41905

[94] Levy, Frank. "Computers and Populism: Artificial Intelligence, Jobs, and Politics in the near Term." *Oxford Review of Economic Policy*, vol. 34, no. 3, July 2018, pp. 393–417. doi:10.1093/oxrep/gry004. https://www.russellsage.org/sites/default/files/gry004.pdf

[95] GE Minds + Machines (November 15, 2016) www.youtube.com/watch?v=OYn9ZtpWCUw

[96] *Why Robots Won't Take over the World.* https://phys.org/news/2018-04-robots-wont-world.html

[97] Freedman, David H. "What Will It Take for IBM's Watson Technology to Stop Being a Dud in Health Care?" *MIT Tech Review* www.technologyreview.com/s/607965/a-reality-check-for-ibms-ai-ambitions/

[98] *Minds and Machines Postponed / Canceled - Post Regarding GE Digital Layoffs.* https://www.thelayoff.com/t/VtbKYAf

[99] Salmon, Felix. "IBM's Watson Was Supposed to Change the Way We Treat Cancer. Here's What Happened Instead." *Slate Magazine*, 18 Aug 2018. https://slate.com/business/2018/08/ibms-watson-how-the-ai-project-to-improve-cancer-treatment-went-wrong.html

[100] Perez, Carlos E. "Why We Should Be Deeply Suspicious of BackPropagation." *Medium*, 13 Oct 2017. https://medium.com/intuitionmachine/the-deeply-suspicious-nature-of-backpropagation-9bed5e2b085e

[101] Datta, Shoumen. *Intelligence in Artificial Intelligence*. Oct 2016. https://arxiv.org/abs/1610.07862

[102] "There's No Good Reason to Trust Blockchain Technology." *Wired*. https://www.wired.com/story/theres-no-good-reason-to-trust-blockchain-technology/

[103] Stinchcombe, Kai. "Blockchain Is Not Only Crappy Technology but a Bad Vision for the Future." *Medium*. 9 April 2018. https://medium.com/@kaistinchcombe/decentralized-and-trustless-crypto-paradise-is-actually-a-medieval-hellhole-c1ca122efdec

[104] "RuuviLab - IOTA Masked Authentication Messaging." *RuuviLab*. https://lab.ruuvi.com/iota

[105] Andreas Kamilaris, Agusti Fonts and Francesc X. Prenafeta-Boldú (2019) The Rise of Blockchain Tech in Agriculture and Food Supply Chains. https://arxiv.org/ftp/arxiv/papers/1908/1908.07391.pdf

[106] Press, Gil. "Big Data Is Dead. Long Live Big Data AI." *Forbes*. www.forbes.com/sites/gilpress/2019/07/01/big-data-is-dead-long-live-big-data-ai/#5a262cf71b05

[107] Walsh, Mary Williams, and Emily Flitter. "McKinsey Faces Criminal Inquiry Over Bankruptcy Case Conduct." *The New York Times*, 8 Nov 2019. https://www.nytimes.com/2019/11/08/business/mckinsey-criminal-investigation-bankruptcy.html

[108] Black, Edwin. *IBM and the Holocaust: The Strategic Alliance between Nazi Germany and America's Most Powerful Corporation.* 1st ed, Crown Publishers, 2001. http://posoh.ru/book/htm/ibm.pdf

[109] Allison, Simon. "Huawei's Pitch to African Mayors: 'Our Cameras Will Make You Safe.'" *The M&G Online*. https://mg.co.za/article/2019-11-15-00-our-cameras-will-make-you-safe

[110] 2007 Joint Workshop on High Confidence Medical Devices, Software, and Systems and Medical Device Plug-and-Play Interoperability: HCMDSS//MD PnP 2007: *Improving Patient Safety through Medical Device Interoperability and High Confidence Software*: Proceedings: 25-27 June 2007, Cambridge, MA. https://nam.edu/wp-content/uploads/2018/02/3.1-Goldman-Jan-2018-002.pdf

[111] Hatcliff, John et al (2011) Medical Application Platforms – Rationale, Architectural Principles, and Certification Challenges. www.nitrd.gov/nitrdgroups/images/8/8b/MedicalDeviceInnovationCPS.pdf

[112] Makary, Martin A., and Michael Daniel. "Medical Error—the Third Leading Cause of Death in the US." *BMJ*, vol. 353, May 2016. doi:10.1136/bmj.i2139 https://www.bmj.com/content/353/bmj.i2139

[113] Slight, Sarah Patricia, et al. "Meaningful Use of Electronic Health Records: Experiences From the Field and Future Opportunities." *JMIR Medical Informatics*, vol. 3, no. 3, 2015, p. e30. doi:10.2196/medinform.4457 https://medinform.jmir.org/2015/3/e30/

[114] *Meaningful Use. CDC.* 10 Sept. 2019 https://www.cdc.gov/ehrmeaningfuluse/introduction.html

[115] MIT AUTO-ID LABORATORY https://autoid.mit.edu/about-lab

[116] Shabandri, Bilal, and Piyush Maheshwari. "Enhancing IoT Security and Privacy Using Distributed Ledgers with IOTA and the Tangle." *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, IEEE, 2019, pp. 1069–75. doi:10.1109/SPIN.2019.8711591 https://assets.ctfassets.net/r1dr6vzfxhev/2t4uxvsIqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota1_4_3.pdf

[117] The Coordicide https://files.iota.org/papers/Coordicide_WP.pdf

[118] https://www.iota.org/

[119] *Send IoT Data to the IOTA Tangle with SAP HANA XSA and Analytics Cloud.* https://blogs.sap.com/2019/10/08/send-iot-data-to-the-iota-tangle-with-sap-hana-xsa-and-analytics-cloud/

[120] https://www.dds-foundation.org/

[121] Mark Shepard, Katherine Baicker and Jonathan S. Skinner (2019). *Does One Medicare Fit All? The Economics of Uniform Health Insurance Benefits* in Tax Policy and the Economy, Volume 34, Moffitt. DOI 10.3386/w26472 https://www.nber.org/papers/w26472.pdf

[122] http://bit.ly/Economics-of-Technology

[123] http://bit.ly/COASE5PAPERS

[124] Merelli, Annalisa. "The Inventor of Microfinance Has an Idea for Fixing Capitalism." *Quartz.* https://qz.com/1089266/the-inventor-of-microfinance-has-an-idea-for-fixing-capitalism/

[125] Georgescu, Cristian. "Simulating Micropayments in Local Area Networks." *Procedia - Social and Behavioral Sciences*, vol. 62, Oct. 2012, pp. 30–34. https://doi.org/10.1016/j.sbspro.2012.09.007

[126] Spence, Michael (2011) *The Next Convergence: The Future of Economic Growth in a Multispeed World.* Farrar, Straus and Giroux, 2011. http://pubdocs.worldbank.org/en/515861447787792966/DEC-Lecture-Series-Michael-Spence-Presentation.pdf

[127] Tabassi, Elham, et al. *A Taxonomy and Terminology of Adversarial Machine Learning.* NIST IR 8269-draft. National Institute of Standards and Technology, October 2019. doi:10.6028/NIST.IR.8269-draft https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf

[128] Duddu, Vasisht. "A Survey of Adversarial Machine Learning in Cyber Warfare." *Defence Science Journal*, vol. 68, no. 4, June 2018, p. 356. doi:10.14429/dsj.68.12371

[129] Victoria Morgan, Lisseth Casso-Hartman, David Bahamon-Pinzon, Kelli McCourt, Robert G. Hjort, Sahar Bahramzadeh, Irene Velez-Torres, Eric McLamore, Carmen Gomes, Evangelyn C. Alocilja, Nirajan Bhusal, Sunaina Shrestha, Nisha Pote, Ruben Kenny Briceno, Shoumen Palit Austin Datta, and Diana C. Vanegas (2020) *Sensor-as-a-Service: Convergence of Sensor Analytic Point Solutions (SNAPS) and Pay-A-Penny-Per-Use (PAPPU) Paradigm as a Catalyst for Democratization of Healthcare in Underserved Communities.* Diagnostics 2020, 10, 22; doi:10.3390/diagnostics10010022 ▪ MIT Library SNAPS - https://dspace.mit.edu/handle/1721.1/123983 ▪ https://dspace.mit.edu/handle/1721.1/111021

[130] Datta, Shoumen et al (2004) *ADAPTIVE VALUE NETWORKS: Convergence of Emerging Tools, Technologies and Standards as Catalytic Drivers* in Chang, Yoon Seok, et al., editors. Evolution of Supply Chain Management: Symbiosis of Adaptive Value Networks and ICT. Kluwer Academic Publishers, 2004. https://dspace.mit.edu/handle/1721.1/41908

[131] Silvestro, Rhian, and Paola Lustrato. "Integrating Financial and Physical Supply Chains: The Role of Banks in Enabling Supply Chain Integration." *International Journal of Operations & Production Management*, vol. 34, no. 3, Jan. 2014, pp. 298–324. doi:10.1108/IJOPM-04-2012-0131. https://www.emerald.com/insight/content/doi/10.1108/IJOPM-04-2012-0131/full/html

[132] Banerjee, Abhijit V., and Esther Duflo. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. 1st ed, PublicAffairs, 2011. https://warwick.ac.uk/about/london/study/warwick-summer-school/courses/macroeconomics/poor_economics.pdf

[133] Kamenetz, Anya. "Esther Duflo Bribes India's Poor To Health." *Fast Company*, 8 August 2011. https://www.fastcompany.com/1768537/esther-duflo-bribes-indias-poor-health

[134] Lie, Robert. *Robertlie/Dht11-Raspi3*. 2018. 2019. *GitHub* https://github.com/robertlie/dht11-raspi3

[135] Handy, Paul. "Introducing Masked Authenticated Messaging." *Medium*, 9 April 2018. https://blog.iota.org/introducing-masked-authenticated-messaging-e55c1822d50e

[136] Ekaterina D. Kazimirova (2017) Human-Centric Internet of Things. Problems and Challenges https://www.researchgate.net/publication/319059870_Human-Centric_Internet_of_Things_Problems_and_Challenges

[137] Calderon, Marco A., et al. "A More Human-Centric Internet of Things with Temporal and Spatial Context." *Procedia Computer Science*, vol. 83, 2016, pp. 553–559 doi:10.1016/j.procs.2016.04.263 https://www.sciencedirect.com/science/article/pii/S1877050916302964

[138] McLamore, E.S., S.P.A. Datta, V. Morgan, N. Cavallaro, G. Kiker, D.M. Jenkins, Y. Rong, C. Gomes, J. Claussen, D. Vanegas, E.C. Alocilja (2019) SNAPS: Sensor Analytics Point Solutions for Detection and Decision Support. *Sensors*, vol. 19, no. 22, p. 4935 ▪ https://www.mdpi.com/1424-8220/19/22/4935/pdf

[139] Artificial Intelligence - Intelligent Agents (2017) https://courses.edx.org/asset-v1:ColumbiaX+CSMM.101x+1T2017+type@asset+block@AI_edx_intelligent_agents_new__1_.pdf

[140] Sutton, Richard S., and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Second edition, MIT Press, 2018. https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

[141] "The Sensor-Based Economy." *Wired*, January 2017. https://www.wired.com/brandlab/2017/01/sensor-based-economy/

142 Pont, Simon, editor. *Digital State: How the Internet Is Changing Everything.* Kogan Page, 2013.

143 Leading the IoT: Gartner Insights on How to Lead in a Connected World

https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf

144 Dave Evans (2011) The Internet of Things: How the Next Evolution of the Internet Is Changing Everything. https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

145 David Puglia. "Are Enterprises Ready for Billions of Devices to Join the Internet?" *Wired*, December 2014. https://www.wired.com/insights/2014/12/enterprises-billions-of-devices-internet/

146 "Gartner Says 5.8 Billion Enterprise and Automotive IoT Endpoints Will Be in Use in 2020." https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io

147 www.cisco.com/c/dam/en/us/products/collateral/se/internet-of-things/at-a-glance-c45-731471.pdf

148 *The Internet of Things: Sizing up the Opportunity. McKinsey.*

https://www.mckinsey.com/industries/semiconductors/our-insights/the-internet-of-things-sizing-up-the-opportunity

149 "IoT Overview Handbook: 2019 Background Primer on The Topics & Technologies Driving the Internet of Things." *Postscapes.* https://www.postscapes.com/iot/

150 Martin Fleming, Wyatt Clarke, Subhro Das, Phai Phongthiengtham, and Prabhat Reddy (2019) The Future of Work: How New Technologies Are Transforming Tasks (October 31, 2019) https://mitibmwatsonailab.mit.edu/research/publications/paper/download/The-Future-of-Work-How-New-Technologies-Are-Transforming-Tasks.pdf

151 Andrew Ng "Why AI Is the New Electricity" https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity

152 Cheekiralla, Sivaram, and Daniel W. Engels. "An IPv6-Based Identification Scheme." *2006 IEEE International Conference on Communications*, vol. 1, 2006, pp. 281–86. doi:10.1109/ICC.2006.254741 https://ieeexplore.ieee.org/document/4024131

153 Stallings, W. "IPv6: The New Internet Protocol." *IEEE Communications Magazine*, vol. 34, no. 7, July 1996, pp. 96–108. doi:10.1109/35.526895. https://ieeexplore.ieee.org/document/526895

154 "Difference Between IPv4 and IPv6" *Tech Differences*, 4 August 2017.

https://techdifferences.com/difference-between-ipv4-and-ipv6.html

155 Datta, Shoumen Palit Austin "An Unified Theory of Relativistic Identification of Information in the Systems Age: Proposed Convergence of Unique Identification with Syntax and Semantics through Internet Protocol Version 6 (IPv6)." *International Journal of Advanced Logistics*, vol. 1, no. 1, July 2012, pp. 66–82. doi:10.1080/2287108X.2012.11006070. https://dspace.mit.edu/handle/1721.1/41902

156 Datta, Shoumen. "Mobile eVote as an IPv6 App." 23 May 2011.

https://shoumendatta.wordpress.com/2011/05/23/mobile-e-vote-ipv6-app/.

https://dspace.mit.edu/bitstream/handle/1721.1/41902/IPv6%20Apps%20and%20SaaS.pdf?sequence=11&isAllowed=y

157 https://bentley.umich.edu/elecrec/d/duderstadt/Speeches/JJDS6/jjd1341.pdf

158 Figueroa, Anthony. "Data Demystified — DIKW Model." *Medium*, 24 May 2019. https://towardsdatascience.com/rootstrap-dikw-model-32cef9ae6dfb

159 William S. Angerman (2004) Coming Full Circle With Boyd's OODA Loop Ideas: An Analysis Of Innovation Diffusion And Evolution (Thesis). https://apps.dtic.mil/dtic/tr/fulltext/u2/a425228.pdf

160 Castanedo, Federico. "A Review of Data Fusion Techniques." *The Scientific World Journal*, 2013, doi:10.1155/2013/704504. http://downloads.hindawi.com/journals/tswj/2013/704504.pdf

161 https://www.wur.nl/en/Education-Programmes/wageningen-academy-1/What-we-offer-you/Courses/show-1/Course-Towards-Data-driven-Agri-Food-Business-1.htm

162 *CS 540 Lecture Notes: Intelligent Agents.* http://pages.cs.wisc.edu/~dyer/cs540/notes/agents.html

163 Maes, Pattie. "Intelligent Software." *Proceedings of the 2nd International Conference on Intelligent User Interfaces - IUI '97*, ACM Press, 1997, pp. 41–43 ▪ doi:10.1145/238218.238283

164 Paolucci, Massimo, and Roberto Sacile. *Agent-Based Manufacturing and Control Systems: New Agile Manufacturing Solutions for Achieving Peak Performance.* CRC Press, 2005. http://jmvidal.cse.sc.edu/library/paolucci05a.pdf

165 Datta, S (2017) Cybersecurity: Agents based Approach? https://dspace.mit.edu/handle/1721.1/107988

166 White, Franklin E. (1997) Data Fusion Group. https://apps.dtic.mil/dtic/tr/fulltext/u2/a394662.pdf

167 Rafferty, Ellen R. S., et al. "Seeking the Optimal Schedule for Chickenpox Vaccination in Canada: Using an Agent-Based Model to Explore the Impact of Dose Timing, Coverage and Waning of Immunity on Disease Outcomes." *Vaccine*, November 2019. doi:10.1016/j.vaccine.2019.10.065.

168 O. Lassila and R.R. Swick. Resource Description Framework (RDF) Model and Syntax Specification. W3C working draft, February 1999. www.w3.org/TR/REC-rdf-syntax/

169 T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. Scientific American, March 2001.

170 Shi, Longxiang, et al. "Semantic Health Knowledge Graph: Semantic Integration of Heterogeneous Medical Knowledge and Services." *BioMed Research International*, vol. 2017, 2017, pp. 1–12 doi:10.1155/2017/2858423 http://downloads.hindawi.com/journals/bmri/2017/2858423.pdf

171 Ayan Chakraborty, Shiladitya Munshi and Debajyoti Mukhopadhyay (2013) *Searching and Establishment of S-P-O Relationships for Linked RDF Graphs : An Adaptive Approach.* 2013 Int Conf on Cloud & Ubiquitous Comp & Emerging Tech https://arxiv.org/ftp/arxiv/papers/1311/1311.7200.pdf

172 Ayan Chakraborty, Shiladitya Munshi and Debajyoti Mukhopadhyay (2013) *A Proposal for the Characterization of MultiDimensional Inter-relationships of RDF Graphs Based on Set Theoretic Approach.* https://arxiv.org/ftp/arxiv/papers/1312/1312.0001.pdf

173 Buneman, Peter. "A Characterisation of Rigid Circuit Graphs." *Discrete Mathematics*, vol. 9, no. 3, Sept. 1974, pp. 205–212 doi:10.1016/0012-365X(74)90002-8. http://homepages.inf.ed.ac.uk/opb/homepagefiles/phylogeny-scans/rigidcircuitgraphs.pdf

174 Sharma, Chandan, and Roopak Sinha. (2019) "A Schema-First Formalism for Labeled Property Graph Databases: Enabling Structured Data Loading and Analytics." *Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies - BDCAT '19*, ACM Press, 2019, pp. 71–80 doi:10.1145/3365109.3368782

175 Wang, Da-Wei, et al. (2019) "CK-Modes Clustering Algorithm Based on Node Cohesion in Labeled Property Graph." *J of Computer Science and Technology*, vol. 34, no. 5, September 2019, pp. 1152-1166 doi:10.1007/s11390-019-1966-0

176 https://aibusiness.com/ending-the-rdf-vs-property-graph-debate-with-rdf/

177 Tarjan, Robert. "Depth-First Search and Linear Graph Algorithms." *SIAM Journal on Computing*, vol. 1, no. 2, June 1972, pp. 146–160 doi:10.1137/0201010.

178 Fleischer, Rudolf, and Gerhard Trippen. (2003) "Experimental Studies of Graph Traversal Algorithms." *Experimental and Efficient Algorithms*, edited by Klaus Jansen et al., vol. 2647, Springer Berlin Heidelberg, 2003, pp. 120–133 doi:10.1007/3-540-44867-5_10

179 Noy, Natasha, et al. "Industry-Scale Knowledge Graphs: Lessons and Challenges." *Com of the ACM*, vol. 62, no. 8, July 2019, pp. 36–43 doi:10.1145/3331166 https://queue.acm.org/detail.cfm?id=3332266

180 https://tech.ebayinc.com/engineering/akutan-a-distributed-knowledge-graph-store/

181 https://www.w3.org/TR/owl-guide/

182 Sengupta K., Hitzler P. (2014) Web Ontology Language (OWL). In: Alhajj, Reda, and Jon Rokne, eds. *Encyclopedia of Social Network Analysis & Mining*. Springer, NY. doi:10.1007/978-1-4614-6170-8

183 http://vowl.visualdataweb.org/v2/

184 https://tools.ietf.org/html/rfc3987

185 www.eads-iw.net/web/nfigay

186 Carbone, Lorenzo, et al. *State of Play of Interoperability: Report 2016.* 2017. *Open WorldCat.* http://dx.publications.europa.eu/10.2799/969314 https://ec.europa.eu/isa2/sites/isa/files/docs/publications/report_2016_rev9_single_pages.pdf

187 "Blind Men and the Elephant." www.allaboutphilosophy.org/blind-men-and-the-elephant.htm

188 Bohr, N. (1950) "On the Notions of Causality and Complementarity." *Science*, vol. 111, no. 2873, January 1950, pages 51–54 doi:10.1126/science.111.2873.51

189 Morrison, Alan (2019) *Is data science/machine learning/AI overhyped right now?* www.quora.com/Is-data-science-machine-learning-AI-overhyped-right-now/answer/Alan-Morrison

190 Fuller, Aidan, et al. "Digital Twin: Enabling Technology, Challenges and Open Research." Oct. 2019. http://arxiv.org/abs/1911.01276 ▪ https://arxiv.org/ftp/arxiv/papers/1911/1911.01276.pdf

191 https://newsstand.joomag.com/en/iic-journal-of-innovation-12th-edition/0994713001573661267

192 https://www.forbes.com/sites/bernardmarr/2017/03/06/what-is-digital-twin-technology-and-why-is-it-so-important

193 Pareto, Vilfredo. *Manual of political economy*. Translated by A. M. Kelley, 1971. MIT Press.

194 Eric S. McLamore, R. Huffaker, Matthew Shupler, Katelyn Ward, Shoumen Palit Austin Datta, M. Katherine Banks, Giorgio Casaburi, Joany Babilonia, Jamie S. Foster (2019) "Digital Proxy of a Bio-Reactor (DIYBOT) Combines Sensor Data and Data Analytics for Wastewater Treatment and Wastewater Management Systems." *(Nature Scientific Reports, in press)* Draft copy of "DIYBOT" available from MIT Libraries https://dspace.mit.edu/handle/1721.1/123983

[195] Hernlund, E., et al. "Osteoporosis in the European Union: Medical Management, Epidemiology and Economic Burden: A Report Prepared in Collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA)." *Archives of Osteoporosis*, vol. 8, no. 1–2, December 2013, p. 136. doi:10.1007/s11657-013-0136-1
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3880487/pdf/11657_2013_Article_136.pdf

[196] Mithal, Ambrish, Dhingra, Vibha and Lau, Edith (2009) .The Asian Audit Epidemiology, costs and burden of osteoporosis in Asia 2009. http://www.iofbonehealth.org/

[197] www.iofbonehealth.org/sites/default/files/PDFs/Audit%20Eastern%20Europe_Central%20Asia/Russian_Audit-Bulgaria.pdf

[198] www.iofbonehealth.org/sites/default/files/PDFs/Audit%20Asia/Asian_regional_audit_Indonesia.pdf

[199] www.iofbonehealth.org/sites/default/files/PDFs/Audit%20Asia/Asian_regional_audit_India.pdf

[200] https://www.iofbonehealth.org/facts-statistics

[201] Global Bone Density Test Market Grows Substantially by 2023, Asserts MRFR Unleashing the Forecast for 2017-2023. *Reuters*. www.reuters.com/brandfeatures/venture-capital/article?id=57229

[202] El Maghraoui, A., et al. "Bone Mineral Density of the Spine and Femur in a Group of Healthy Moroccan Men." *Bone*, vol. 44, no. 5, May 2009, pp. 965–969 doi:10.1016/j.bone.2008.12.025

[203] Angell, Marcia. (2005) *The Truth about the Drug Companies: How They Deceive Us and What to Do about It*. Random House Trade Paperbacks, 2005.
https://cyber.harvard.edu/cyberlaw2005/sites/cyberlaw2005/images/NYReviewBooksAngell.pdf

[204] https://www.keele.ac.uk/pharmacy-bioengineering/ourpeople/jamesnolan/

[205] https://people.eecs.berkeley.edu/~pabbeel/

[206] http://bit.ly/Farm-IoT-Ranveer

[207] https://www.theaaih.org/

[208] Gutterman, D.J. (2009) Silent Myocardial Ischemia. May 2009. Circulation Journal 2009 **73** 785–797.
https://www.jstage.jst.go.jp/article/circj/73/5/73_CJ-08-1209/_pdf

[209] https://github.com/alirezadir/Production-Level-Deep-Learning

[210] Hannah Ritchie (2018) *What do people die from?* https://ourworldindata.org/what-does-the-world-die-from

[211] https://www.who.int/nmh/publications/ncd_report_chapter1.pdf

[212] Institute for Health Metrics and Evaluation, Human Development Network, The World Bank. The Global Burden of Disease: Generating Evidence, Guiding Policy — Sub-Saharan Africa Regional Edition. Seattle, WA: IHME, 2013.
http://documents.worldbank.org/curated/en/831161468191672519/pdf/808520PUB0ENGL0Box037982 0B00PUBLIC0.pdf

[213] Global health risks: mortality and burden of disease attributable to selected major risks. World Health Organization (2009). ISBN 978 92 4 156387 1
https://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf

[214] https://www.who.int/pmnch/media/press_materials/fs/fs_mdg4_childmortality/en/

215 https://economictimes.indiatimes.com/magazines/panache/jack-dorsey-has-fallen-in-love-with-africa-plans-to-shift-there-for-3-6-months-next-year/articleshow/72275949.cms

216 The concept of "bit dribbling" may not be new but may be attributed to Neil Gershenfeld (MIT) circa 1999 (*personal communication*). The notion is that very small amounts of data and information (bits) transmitted (dribbled) at the right time, are often enough to serve 80% of cases, in several circumstances.

217 Parker, Martin (2018) Shut Down the Business School.

https://www.theguardian.com/news/2018/apr/27/bulldoze-the-business-school

218 Eid, Mohamad A., et al. "A Novel Eye-Gaze-Controlled Wheelchair System for Navigating Unknown Environments: Case Study With a Person With ALS." *IEEE Access*, vol. 4, 2016, pages 558 - 573. doi:10.1109/ACCESS.2016.2520093.

219 *Robots for Lonely Hearts – Asian Robotics Review*. https://asianroboticsreview.com/home29-html

220 Oppenheimer, Andres, and Ezra E. Fitz. *The Robots Are Coming! The Future of Jobs in the Age of Automation*. Vintage Books, a division of Penguin Random House, 2019. ISBN-13 978-0525565000

221 Deming, David. "The Robots Are Coming. Prepare for Trouble." *The New York Times*, 30 January 2020. https://www.nytimes.com/2020/01/30/business/artificial-intelligence-robots-retail.html

222 http://www.philipkdickfans.com/mirror/websites/pkdweb/short_stories/Autofac.htm

223 https://www.bloomberg.com/news/features/2017-10-18/this-company-s-robots-are-making-everything-and-reshaping-the-world

224 Ford, Martin. *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books, 2016.

225 Pugliano, John. *The Robots Are Coming: A Human's Survival Guide to Profiting in the Age of Automation*. Ulysses Press, 2017.

226 The Dutch Safety Board (2009) *Crashed during approach, Boeing 737-800, near Amsterdam Schiphol Airport, 25 February 2009.* https://catsr.vse.gmu.edu/SYST460/TA1951_AccidentReport.pdf

227 Pacaux-Lemoine, Marie-Pierre, and Frank Flemisch. "Layers of Shared and Cooperative Control, Assistance and Automation." *IFAC-PapersOnLine*, vol. 49, no. 19, 2016, pp. 159 - 164 doi:10.1016/j.ifacol.2016.10.479

228 Wickens, Christopher D., et al. "Stages and Levels of Automation: An Integrated Meta-Analysis." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 54, no. 4, Sept. 2010, pp. 389 - 393 doi:10.1177/154193121005400425

229 Hamby, Chris. "How Boeing's Responsibility in a Deadly Crash 'Got Buried.'" *The New York Times*, 20 January 2020. https://www.nytimes.com/2020/01/20/business/boeing-737-accidents.html

230 Hamby, C. and Moses, C. "Boeing Refuses to Cooperate With New Inquiry Into Deadly Crash." *The New York Times*, 6 February 2020. www.nytimes.com/2020/02/06/business/boeing-737-inquiry.html

231 Solow, Robert M. (1957) "Technical Change and the Aggregate Production Function." *The Review of Economics and Statistics*, vol. 39, no. 3, Aug. 1957, p. 312. doi:10.2307/1926047

http://www.piketty.pse.ens.fr/files/Solow1957.pdf

232 http://ghdx.healthdata.org/

233 https://www.oxfordeconomics.com/recent-releases/how-robots-change-the-world

[234] https://apm.byu.edu/prism/index.php/Members/JohnHedengren

[235] https://apmonitor.com/do/index.php/Main/ShortCourse

[236] Osgood, Nathaniel. (2019) Systems Data Science.
https://www.youtube.com/watch?v=CPUOyqs9G3Q&feature=youtu.be

[237] Bayes, Thomas. (1763) *An essay towards solving a problem in the doctrine of chances.* Philosophical Transactions of the Royal Society of London 53(0):370-418, 1763 http://www.rssb.be/bsn57/bsn57-6.pdf

[238] Exceptions prove the rule. Not all Indians espouse the "body shop" mantra. Global business and tech leaders (2020) of Indian origin include Mrs Jayashree Ullal (CEO, Arista Networks), Mr Arvind Krishna (CEO, IBM, incoming), Mr Sundar Pichai (CEO, Alphabet/Google), Mr Satya Nadella (CEO, Microsoft), Mr Shantanu Narayen (CEO, Adobe), Mr Rajeev Suri (CEO, Nokia, outgoing), Mr V K Narasimhan (CEO, Novartis) and Mr Ajaypal Singh Banga (CEO, MasterCard). The market cap of these 8 companies (~$3T) may be 20% of the total market cap of the top 50 US companies (~$15T). www.iweblists.com/us

[239] Sabrie, Gilles (2019) *Behind the Rise of China's Facial-Recognition Giants* (09.03.2019)
https://www.wired.com/story/behind-rise-chinas-facial-recognition-giants/

[240] https://www.purdue.edu/rosehub/

[241] Navi Radjou & Jaideep Prabhu (2016) *Frugal Innovation: How to do More With Less.*
http://naviradjou.com/book/frugal-innovation-how-to-do-more-with-less/

[242] Creighton, Jean (2019) 5 Moon-landing innovations that changed life on Earth
https://theconversation.com/5-moon-landing-innovations-that-changed-life-on-earth-102700

[243] https://www.inc.com/bill-murphy-jr/27-innovations-we-use-constantly-but-that-you-probably-didnt-know-were-from-nasa-space-program.html

[244] https://spinoff.nasa.gov/

[245] Marcus, Gary (2019) *An Epidemic of AI Misinformation.*
https://thegradient.pub/an-epidemic-of-ai-misinformation/

[246] Vanegas, D. C., L. Patiño, C. Mendez, D. A. Oliveira, A. M. Torres, C. L. Gomes, and E. S. McLamore. "Laser Scribed Graphene Biosensor for Detection of Biogenic Amines in Food Samples Using Locally Sourced Materials." *Biosensors*, vol. 8, no. 2, Apr. 2018, p. 42. doi:10.3390/bios8020042
https://www.mdpi.com/2079-6374/8/2/42

[247] Holmes, E. A., et al. "Applications of Time-Series Analysis to Mood Fluctuations in Bipolar Disorder to Promote Treatment Innovation: A Case Series." *Translational Psychiatry*, vol. 6, no. 1, Jan. 2016, pp. e720–e720. doi:10.1038/tp.2015.207
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5068881/pdf/tp2015207a.pdf

[248] Trautmann, Sebastian, et al. *"The Economic Costs of Mental Disorders: Do Our Societies React Appropriately to the Burden of Mental Disorders?" EMBO Reports*, vol. 17, no. 9, Sept. 2016, pp. 1245–1249. doi:10.15252/embr.201642951
www.ncbi.nlm.nih.gov/pmc/articles/PMC5007565/pdf/EMBR-17-1245.pdf

[249] Drevets, Wayne C., et al. "Subgenual Prefrontal Cortex Abnormalities in Mood Disorders." *Nature*, vol. 386, no. 6627, Apr. 1997, pp. 824–827 doi:10.1038/386824a0

[250] P. Llamocca, A. Junestrand, M. Cukic, D. Urgelés, V. López. (2018) *Data source analysis in mood disorder research.* XVIII Proceedings of the XVIII Conference of the Spanish Association for Artificial Intelligence, CAEPIA, ISBN: 978-84-09-05643-9 F. Herrera et al. (Eds.), pp. 893-900, Granada, Spain.

[251] P. Llamocca, D. Urgelés, M. Cukic, V. Lopez. (2019) *Bip4Cast: Some advances in mood disorders data analysis.* Proceedings of the 1st International Alan Turing Conference on Decision Support and Recommender Systems, London 2019.

[252] Raymond Perrault, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz, Terah Lyons, James Manyika, Saurabh Mishra, and Juan Carlos Niebles, "The AI Index 2019 Annual Report", AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, December 2019. https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf

[253] Chen, Yu-Hsin, et al. "Eyeriss v2: A Flexible Accelerator for Emerging Deep Neural Networks on Mobile Devices." May 2019. http://arxiv.org/abs/1807.07928 ▪ https://arxiv.org/pdf/1807.07928.pdf https://www.rle.mit.edu/eems/wp-content/uploads/2019/04/2019_jetcas_eyerissv2.pdf

[254] Jo, Eun Seo, and Timnit Gebru. "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning." Dec. 2019 doi:10.1145/3351095.3372829 https://arxiv.org/pdf/1912.10389.pdf

[255] Tembon, Mercy Miyang, and Lucia Fort, editors. *Girl's Education in the 21st Century: Gender Equality, Empowerment and Growth*. The World Bank, 2008. doi:10.1596/978-0-8213-7474-0

[256] https://www.nytimes.com/2006/10/14/world/asia/14nobel.html

[257] Cohn, Alain, et al. "Civic Honesty around the Globe." *Science*, vol. 365, no. 6448, July 2019, pp. 70–73. doi:10.1126/science.aau8712

[258] Grosch, Kerstin; Rau, Holger (2017) : Gender differences in honesty: The role of social value orientation. Discussion Papers, No. 308, University of Göttingen, Center for European, Governance and Economic Development Research (CEGE), Göttingen, Germany. https://www.econstor.eu/bitstream/10419/156226/1/882555200.pdf

[259] Sen, Amartya. *The Idea of Justice*. Harvard University Press, 2009. https://dutraeconomicus.files.wordpress.com/2014/02/amartya-sen-the-idea-of-justice-2009.pdf

[260] "Artificial Intelligence Makes Bad Medicine Even Worse." *Wired*. www.wired.com, https://www.wired.com/story/artificial-intelligence-makes-bad-medicine-even-worse/

[261] Datta, Shoumen (2008) Arm Chair Essays in Energy. https://dspace.mit.edu/handle/1721.1/45512

[262] Donald Mackenzie (1877) The Flooding of the Sahara: An Account of the Proposed Plan for Opening Central Africa to Commerce and Civilization. 1877. S. Low, Marston, Searle, & Rivington, Publishers. https://ia902205.us.archive.org/23/items/floodingsaharaa01mackgoog/floodingsaharaa01mackgoog.pdf

[263] https://www.jstor.org/stable/1761255?seq=1#metadata_info_tab_contents

[264] https://www.nature.com/articles/019509a0.pdf

[265] Rothman, Joshua. *The Equality Conundrum* 13 January 2020. www.newyorker.com/magazine/annals-of-inquiry ◆ https://www.newyorker.com/magazine/2020/01/13/the-equality-conundrum

[266] https://www.mikespastry.com/

[267] https://www.modernpastry.com/

[268] Ilić, Suzana, et al. (2018) "Deep Contextualized Word Representations for Detecting Sarcasm and Irony." *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Association for Computational Linguistics, 2018, pp. 2–7 doi:10.18653/v1/W18-6202 https://arxiv.org/pdf/1809.09795.pdf

[269] Datta, Shoumen (2008) Convergence of Bio, Info, Nano, Eco: Global Public Goods and Economic Growth. https://dspace.mit.edu/handle/1721.1/41909

[270] Steinmueller, W. Edward, 2010. "Economics of Technology Policy" Handbook of the Economics of Innovation, in: Bronwyn H. Hall & Nathan Rosenberg (ed.), Handbook of the Economics of Innovation, edition 1, volume 2, pages 1181-1218, Elsevier ▪ https://doi.org/10.1016/S0169-7218(10)02012-5

[271] Singh, Lakhveer, et al. "Bioelectrofuel Synthesis by Nanoenzymes: Novel Alternatives to Conventional Enzymes." *Trends in Biotechnology*, Jan. 2020, p. S0167779919303129 doi:10.1016/j.tibtech.2019.12.017

[272] L. Wang, Y. Chen, F. Long, L. Singh, S. Trujillo, X. Xiao, H. Liu (2020) *Breaking the Loop: Tackling Homoacetogenesis by Chloroform to Halt Hydrogen Production-Consumption Loop in Single Chamber Microbial Electrolysis Cells.* Chemical Engineering Journal. https://doi.org/10.1016/j.cej. 2020.124436

[273] https://patents.justia.com/inventor/andrew-e-fano

[274] https://patents.justia.com/patent/10095981

[275] https://claimparse.com/patent.php?patent_num=10275721

[276] https://www.ibm.com/downloads/cas/KMPVGB4W

[277] https://mcgovern.mit.edu/profile/tomaso-poggio/

[278] Faro, Scott H., and Feroze B. Mohamed, editors. *Functional MRI: Basic Principles and Clinical Applications*. Springer, 2006. ISBN 978-0-387-23046-7

[279] https://towardsdatascience.com/bert-roberta-distilbert-xlnet-which-one-to-use-3d5ab82ba5f8

[280] https://cloud.google.com/tpu/

[281] https://blog.floydhub.com/when-the-best-nlp-model-is-not-the-best-choice/

[282] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer (2018) *Deep contextualized word representations.* https://arxiv.org/pdf/1802.05365.pdf

[283] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018) *BERT: pre-training of deep bidirectional transformers for language understanding.* https://arxiv.org/pdf/1810.04805.pdf

[284] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma and Radu Soricut (2020) *Albert: A Lite Bert for Self-Supervised Learning of Language Representations.* 8th International Conference on Learning Representations (2020) https://openreview.net/pdf?id=H1eA7AEtvS

[285] Sun, Cong, et al. "A Deep Learning Approach With Deep Contextualized Word Representations for Chemical–Protein Interaction Extraction From Biomedical Literature." *IEEE Access*, vol. 7, 2019, pp. 151034–46 doi:10.1109/ACCESS.2019.2948155

[286] Vahe Tshitoyan, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder and Anubhav Jain (2019) *Unsupervised word embeddings capture latent knowledge from materials science literature. Nature* **571,** 95–98 (2019) https://doi.org/10.1038/s41586-019-1335-8

[287] Babyak, M. A. (2004) "What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models." *Psychosomatic Med*, vol. 66, no. 3, May 2004, pp 411-21 DOI 10.1097/01.psy.0000127692.23278.a9 https://people.duke.edu/~mababyak/papers/babyakregression.pdf

[288] Tarfa Hamed (2017) *Recursive Feature Addition: A Novel Feature Selection Technique, Including a Proof of Concept in Network Security.* PhD thesis submitted to University of Guelph, Ontario, Canada. https://atrium.lib.uoguelph.ca/xmlui/bitstream/handle/10214/10315/Hamed_Tarfa_201704_PhD.pdf?sequence=1&isAllowed=y

[289] Aditya Nandy, Chenru Duan, Jon Paul Janet, Stefan Gugler, and Heather J. Kulik (2018) "Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry." *Industrial and Engineering Chemistry Research*, volume 57, number 42, October 2018, pages 13973 - 13986 doi:10.1021/acs.iecr.8b04015

[290] Virshup, Aaron M., et al. "Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds." *Journal of the American Chemical Society*, vol. 135, no. 19, May 2013, pp. 7296 - 7303 doi:10.1021/ja401184g

[291] Sam Lemonick (2019) As DFT Matures, Will It Become a Push-Button Technology? *Chemical & Engineering News*, volume 97, issue 35 https://cen.acs.org/physical-chemistry/computational-chemistry/DFT-matures-become-push-button/97/i35

[292] Joan Bruna, Wojciech Zaremba, Arthur Szlam & Yann LeCun (2014) "Spectral Networks and Locally Connected Networks on Graphs." http://arxiv.org/abs/1312.6203 ▪ https://arxiv.org/pdf/1312.6203.pdf

[293] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals and George E. Dahl (2017) *Neural Message Passing for Quantum Chemistry.* https://arxiv.org/pdf/1704.01212.pdf

[294] Kyle Swanson (2019) Message Passing Neural Networks for Molecular Property Prediction. Master's Thesis. EECS, MIT. https://dspace.mit.edu/bitstream/handle/1721.1/123133/1128814048-MIT.pdf?sequence=1&isAllowed=y

[295] http://chemprop.csail.mit.edu/

[296] Lars Ruddigkeit, Ruud van Deursen, Lorenz C. Blum and Jean-Louis Reymond (2012) "Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17." *J of Chemical Information and Modeling*, vol. 52, no. 11, November 2012, pp. 2864–2875 doi:10.1021/ci300415d

[297] Jonathan M. Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M. Donghia, Craig R. MacNair, Shawn French, Lindsey A. Carfrae, Zohar Bloom-Ackerman, Victoria M. Tran, Anush Chiappino-Pepe, Ahmed H. Badran, Ian W. Andrews, Emma J. Chory, George M. Church, Eric D. Brown, Tommi S. Jaakkola, Regina Barzilay and James J. Collins (2020) "A Deep Learning Approach to Antibiotic Discovery." *Cell*, vol. 180, no. 4, February 2020, pp. 688-702.e13 doi:10.1016/j.cell.2020.01.021 ◆ https://www.cell.com/cell/pdf/S0092-8674(20)30102-1.pdf

[298] http://news.mit.edu/2020/artificial-intelligence-identifies-new-antibiotic-0220

[299] Zhang, Xiang, et al. (2019) "Deep Neural Network Hyperparameter Optimization with Orthogonal Array Tuning." July 2019. http://arxiv.org/abs/1907.13359 ◆ https://arxiv.org/pdf/1907.13359.pdf

[300] Lars Kai Hansen and Peter Salamon (1990) *Neural Network Ensembles*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 10, October 1990.
https://pdfs.semanticscholar.org/257d/c8ae2a8353bb2e86c1b7186e7d989fb433d3.pdf

[301] Marcello Pelillo (2014) "Alhazen and the nearest neighbor rule." *Pattern Recognition Letters* 38 (2014) 34–37 http://dx.doi.org/10.1016/j.patrec.2013.10.022

[302] https://thenextweb.com/neural/2020/02/19/study-ai-expert-gary-marcus-explains-how-to-take-ai-to-the-next-level/

[303] Russell, Stuart and Norvig, Peter (2010) *Artificial Intelligence: A Modern Approach*. 3rd ed, Prentice Hall. http://aima.cs.berkeley.edu/

[304] Winston, Patrick Henry and Richard Henry Brown, editors. *Artificial Intelligence, an MIT Perspective*. MIT Press, 1979. https://courses.csail.mit.edu/6.034f/ai3/rest.pdf

[305] Buchanan, Bruce G., and Edward Hance Shortliffe, editors. (1984) *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley. 1984.
http://digilib.stmik-banjarbaru.ac.id/data.bc/2.%20AI/2.%20AI/1984%20Rule-Based%20Expert%20Systems.pdf

[306] Quinlan, J. R., editor. (1987) *Applications of Expert Systems: Based on the Proceedings of the Second Australian Conference*. Turing Institute Press in association with Addison-Wesley Pub. Co, 1987.

[307] Giarratano, Joseph C. and Gary Riley. (1989) *Expert Systems: Principles and Programming*. 1st ed. PWS-Kent Publishing Company, Boston. ISBN 0-87835-335-6

[308] Buchanan, Bruce G. (1989) "Can Machine Learning Offer Anything to Expert Systems?" *Machine Learning*, vol. 4, no. 3–4, December 1989, pages 251–254 doi:10.1007/BF00130712.
https://link.springer.com/content/pdf/10.1007/BF00130712.pdf

[309] Valverde, S. (2016) Major transitions in information technology. Philosophical Transactions of the Royal Society. B 371: 20150450 ◆ http://dx.doi.org/10.1098/rstb.2015.0450
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4958943/pdf/rstb20150450.pdf

[310] Mearian, Lucas. (2017) "CW@50: Data Storage Goes from $1M to 2 Cents per Gigabyte (+video)." *Computerworld*, 23 March 2017 ◆ https://www.computerworld.com/article/3182207/cw50-data-storage-goes-from-1m-to-2-cents-per-gigabyte.html

[311] Supernor, Bill (2018) "Why the Cost of Cloud Computing Is Dropping Dramatically." *App Developer Magazine*, https://appdevelopermagazine.com/why-the-cost-of-cloud-computing-is-dropping-dramatically/

[312] Wolpert, H David (2018). "Why Do Computers Use So Much Energy?" *Scientific American Blog Network* ▪ https://blogs.scientificamerican.com/observations/why-do-computers-use-so-much-energy/

[313] David H Wolpert. (2019) "The stochastic thermodynamics of computation." *Journal of Physics A: Mathematical and Theoretical 2019* https://iopscience.iop.org/article/10.1088/1751-8121/ab0850/pdf

[314] Panesar, S.S., Kliot, M., Parrish, R., Fernandez-Miranda, J., Cagle, Y., Britz, G.W. (2019) Promises and Perils of Artificial Intelligence in Neurosurgery. *Neurosurgery* ▪ https://doi.org/10.1093/neuros/nyz471

---

[315] Emily Hanhauser, Michael S. Bono, Jr., Chintan Vaishnav, A. John Hart, and Rohit Karnik (2020) Solid-Phase Extraction, Preservation, Storage, Transport, and Analysis of Trace Contaminants for Water Quality Monitoring of Heavy Metals, *Environ Science & Tech* (2020). DOI: 10.1021/acs.est.9b04695

[316] Rohit Sharma , Sachin S. Kamble , Angappa Gunasekaran , Vikas Kumar , Anil Kumar (2020) A Systematic Literature Review on Machine Learning Applications for Sustainable Agriculture Supply Chain Performance, Computers and Operations Research. https://doi.org/10.1016/j.cor.2020.104926

[317] Amelie Gyrard, Manas Gaur, Saeedeh Shekarpour, Krishnaprasad Thirunarayan and Amit Sheth (2018) *Personalized Health Knowledge Graph.* Contextualized Knowledge Graph Workshop, International Semantic Web Conference, 2018.
https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=1005&context=aii_fac_pub

[318] "Apple Agrees to Settlement of up to $500 Million from Lawsuit Alleging It Throttled Older Phones." *TechCrunch*, http://social.techcrunch.com/2020/03/02/apple-agrees-to-settlement-of-up-to-500-million-from-lawsuit-alleging-it-throttled-older-phones/

[319] Gini, Corrado (1909) "Concentration and dependency ratios" (in Italian). English translation in *Rivista di Politica Economica* (1997) 87: 769–789

[320]
www.reddit.com/r/MapPorn/comments/88mw3q/open_defecation_around_the_world_2015_960_684/

[321] "Nearly a Billion People Still Defecate Outdoors. Here's Why." *Magazine*, 25 July 2017,
https://www.nationalgeographic.com/magazine/2017/08/toilet-defecate-outdoors-stunting-sanitation/

[322] Sadhu, Bodhisatwa, et al. "The More (Antennas), the Merrier: A Survey of Silicon-Based Mm-Wave Phased Arrays Using Multi-IC Scaling." *IEEE Microwave Magazine*, vol. 20, no. 12, Dec 2019, pp 32–50

[323] Wang, Tengfei and Kang, Jong Woo (2020). *An integrated approach for assessing national e-commerce performance.* Trade, Investment and Innovation Working Paper Series, No. 01/20, ESCAP Trade, Investment and Innovation Division, United Nations (UN). January 2020. Bangkok, Thailand.
https://www.unescap.org/sites/default/files/publications/Working%20Paper%20No.1_2020.pdf

[324] http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf

[325] Propaganda erudite, credible and respectable scientists, are disturbing, devastating and desacralizing. [This statement and this note must be taken with more than just a pinch of salt because it is the personal opinion of the primary author (SD), who may be best described by others as less than academic dust along the infinite corridors of the esteemed institutions with which he may be affiliated.] Time, events and publications suggest that "giants" who are *good* are occasionally hypnotized by the slippery slope of metamorphosis from *good* to self-anointed "God" particles. Knowledge, which was once regarded as an oak tree, and supposed to usher in self-deprecation, modesty and humility, now, frequently suffers from bloating, sufficient to spill over the *black hole* of hubris. For example, a trio of brilliant (see ref 91) male scientists, in the upper latitudes of North America, are acting as *Nostradamus*, stoked by greed and the quest for immortality, buoyed by corporate largesse, exclusively driven by the desire for wealth creation. A group of complicit organizations and ill-informed media are ever ready to quench the drab voices of reason and restraint, in favour of sensationalizing and amplifying this inane *Nostradamus Effect.*

This harms society (due to derelict reports, for example, see reference number 233) and reduces the credibility of august institutions and organizations which appear as pawns for corporate business development (https://knowledgegraphsocialgood.pubpub.org) often under a camouflage of so-called knowledge for social good. https://knowledgegraphsocialgood.pubpub.org/programcomittee

326 Boroush, Mark (2020) National Science Board, National Science Foundation. 2020. Research and Development: U.S. Trends and International Comparisons. Science and Engineering Indicators 2020. NSB-2020-3. Alexandria, VA. https://ncses.nsf.gov/pubs/nsb20203/assets/nsb20203.pdf

327 Cyranoski, David. "China Awaits Controversial Blacklist of 'Poor Quality' Journals." *Nature*, vol. 562, Oct. 2018, pp. 471–72. *www.nature.com*, doi:10.1038/d41586-018-07025-5

328 https://www.thehindu.com/profile/photographers/Prashant-Waydande/

329 "Indian Research Quality Lags Quantity." *Economic Times Blog*, 24 Dec. 2019, https://economictimes.indiatimes.com/blogs/et-editorials/indian-research-quality-lags-quantity/

330 Jaishankar, Dhruva. "The Huge Cost of India's Discrimination Against Women." *The Atlantic*, 18 Mar. 2013. https://www.theatlantic.com/international/archive/2013/03/the-huge-cost-of-indias-discrimination-against-women/274115/

331 Kroll, Barry M. "Writing for Readers: Three Perspectives on Audience." *College Composition and Communication*, vol. 35, no. 2, May 1984, p. 172. doi:10.2307/358094

332 Soto C., R., Robles-Baldenegro, M.E., López, V. and Camalich, J.A. (2017), MQDM: An Iterative Fuzzy Method for Group Decision Making in Structured Social Networks. International Journal of Intelligent Systems **32** pages 17-30 ◆ https://onlinelibrary.wiley.com/doi/full/10.1002/int.21826

333 Schwartz, M. A (2008) Importance of Stupidity in Scientific Research. *Journal of Cell Science*, vol. 121, no. 11, pp. 1771, doi:10.1242/jcs.033340 ▪ https://jcs.biologists.org/content/joces/121/11/1771.full.pdf

334 http://shakespeare.mit.edu/macbeth/macbeth.5.5.html

335 Jean le Rond d'Alembert (1886) Oeuvres de D'Alembert. ISBN-13 9781103953189

336 Gordin, Michael D. (2015) *Scientific Babel: How Science Was Done before and after Global English.* The University of Chicago Press, 2015.

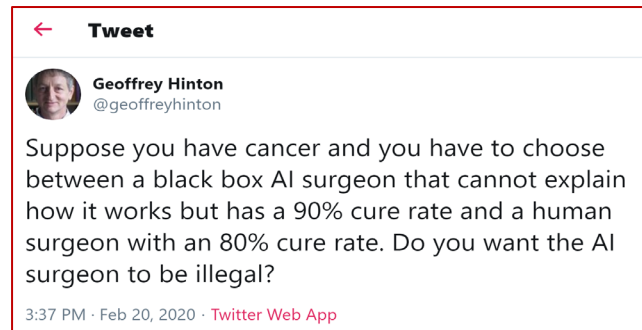337 https://citations.ouest-france.fr/citation-voltaire/medecins-administrent-medicaments-dont-savent-22351.html

Market proponents of AI are individuals who blather about neural networks of which they know little, to solve problems using learning tools which they know less, for the society of human beings of whom they know nothing.

(**Adapted from** "Les médecins administrent des médicaments dont ils savent très peu, à des malades dont ils savent moins, pour guérir des maladies dont ils ne savent rien" – **Voltaire**[337])

(*Doctors are men who prescribe medicines of which they know little, to cure diseases of which they know less, in human beings of whom they know nothing.*)

Tweet

Geoffrey Hinton
@geoffreyhinton

Suppose you have cancer and you have to choose between a black box AI surgeon that cannot explain how it works but has a 90% cure rate and a human surgeon with an 80% cure rate. Do you want the AI surgeon to be illegal?

3:37 PM · Feb 20, 2020 · Twitter Web App

IS THIS A BLANK PAGE? WE SEE AN UNSEEN SENSE OF THE FUTURE IN WHAT APPEARS TO BE A BLANK PAGE.

We are keeping room for disambiguation, discussion, dissent (and imagination, invention, innovation).

Good leadership requires you to surround yourself with people of diverse perspectives who can disagree with you