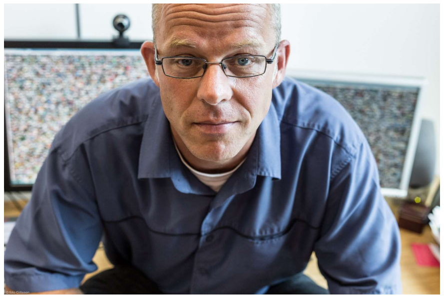
Interactive Learning

Human-Centred Machine Learning, 2024/25

MARCEL WORRING

Who Am I?

- Full professor in Multimedia Analytics
- Leading the MultiX research group



We research **multimedia analytics** techniques for getting the richest information possible from the data through AI algorithms, interactions, and interfaces; surpassing human and machine intelligence for applications and social impact in public health, forensics and law enforcement, cultural heritage, and data-driven business.

Introduction

The goals in AI

What projects did you do in AI research?

What should we aim for in AI?

- Useful AI
  - can be defined as an AI solution that can provide the functions required to satisfy target users' needs in the valid usage scenarios of their work and life.
- Usable AI
  - can be defined as an AI solution that is easy to learn and use via optimal user experience (UX) created by effective Human-Computer Interaction (HCI) design

[Xu2019]

Beyond Algorithms

- Current AI research
  - Is focused on a fully automatic process
  - Is focused on improving the accuracy of models or at least well-defined quantitative metrics
- When striving for useful and usable AI
  - There is so much more than automatic methods that get a good score on one of the metrics

Useful AI is (mostly) Trustworthy AI

- Accuracy
  - How well does the AI system do on new (unseen) data compared to data on which it was trained and tested?
- Robustness
  - How sensitive is the system's outcome to a change in the input?
- Fairness
  - Are the system outcomes unbiased?
- Accountability
  - Who or what is responsible for the system's outcome?
- Transparency
  - Is it clear to an external observer how the system's outcome was produced?
- Explainability
  - Can the system's outcome be justified with an explanation that a human can understand and/or that is meaningful to the end user?
- Ethical
  - Was the data collected in an ethical manner? Will the system's outcome be used in an ethical manner?

[Mosqueira-Rey2022]

Usable AI starts with Data Usability

- Usable
  - usability in the most limited context, i.e., that we can use them because they are the right size, we have the right permissions, their license allows it, they do not contain sensitive information, etc.
- Relevant
  - cover the topic of interest at the right level of detail.
- Quality
  - completeness, provenance, accuracy, cleanliness, consistency of formatting, etc.
- Reusable
  - can be used in different studies where we would incorporate aspects such as that they are easily understandable, that there are different ways of accessing them, that there is a management of the changes produced in the data, as well as a collaborative nature in the data work processes.

[Mosqueira-Rey2022]

Usable AI needs to consider its different users

- ML experts
  - Experts with extensive knowledge in ML techniques.
    - Supervised learning: they select the data, label the data, classify them into training data and testing data, extract the features needed to feed the machine learning algorithm, create the model and refine it if the performance obtained is not optimal, etc.
    - Unsupervised learning: machine learning experts are required to interpret the clusters identified by the model so data can be converted into knowledge.
- Domain experts
  - In many domains, the designers of machine-learning-based systems do not themselves hold the expertise required to create training data. In such projects, the collaboration of domain experts is necessary.
- Data experts
  - A data expert or data scientist is a multi-disciplinary scientist that uses methods, processes and algorithms to extract knowledge from data.

[Mosqueira-Rey2022]

AI and Visual Analytics

- The fundamental parts of an IML system
  - Users, model, data, interface
- Two quotes from the review papers
  - The machine learning perspective
    - "*A classic ML system must have an interface, but it is a passive one; in IML the interface is responsible for the bidirectional feedback between the other three components and for the authors, the interface design is critical to the success of the IML process*" [Mosqueira-Rey2022]
  - The visual analytics perspective
    - "*VA tools have the potential for improved support of ML interpretation, understandability, validation, and refinement through interaction. However, current VA tools and ML components are posing many interesting challenges for future work. Towards addressing these challenges, closer collaboration between ML and visualization researchers is vital*" [Sacha2016]

Interdisciplinary research is essential

- There are no clear definitions (yet)
- Overviews as starting point
- Different from automatic AI not benchmark with “simple” improvement
- Therefore, today we focus on several interrelated review papers

The algorithmic view

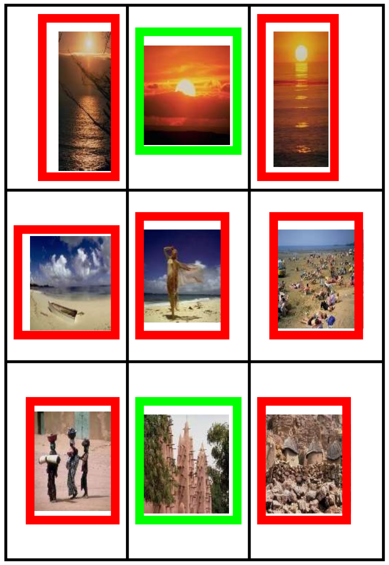
[MOSQUIRA-REY2022]

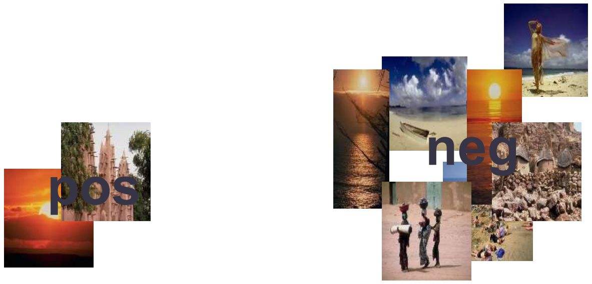
Different ways of learning

- Active learning (AL)
  - the system remains in control of the learning process and treats humans as oracles to annotate unlabeled data.
  - special case is Relevance Feedback where the user is giving feedback on the output
- Interactive machine learning (IML)
  - a closer interaction between users and learning systems, with people interactively supplying information in a more focused, frequent, and incremental way compared to traditional machine learning.
- Machine teaching
  - human domain experts have control over the learning process by delimiting the knowledge that they intend to transfer to the machine learning model.

Traditional ways of employing user interaction

Selection of pos/neg examples

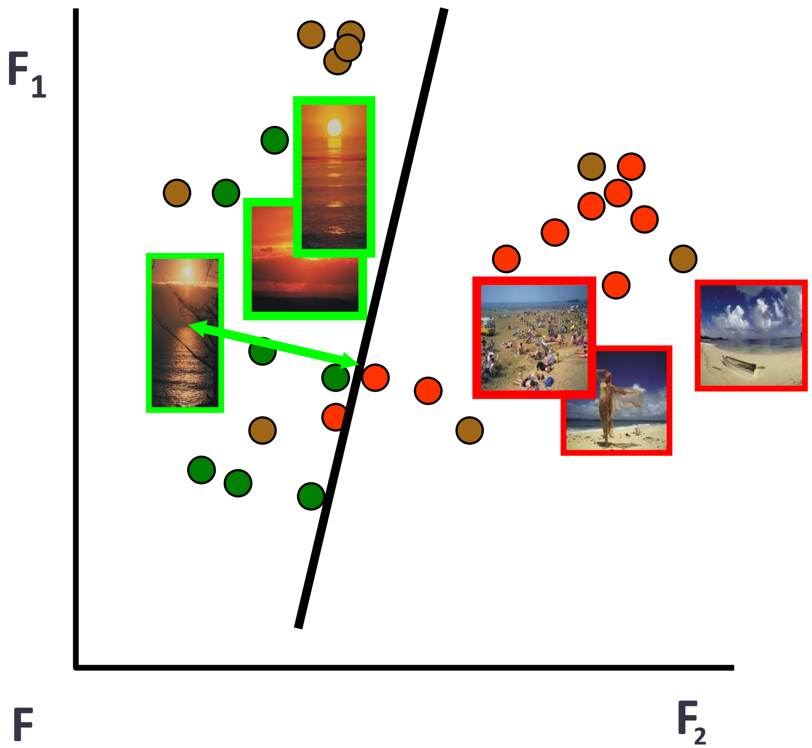




Some elements in the collection are labeled

Many are not

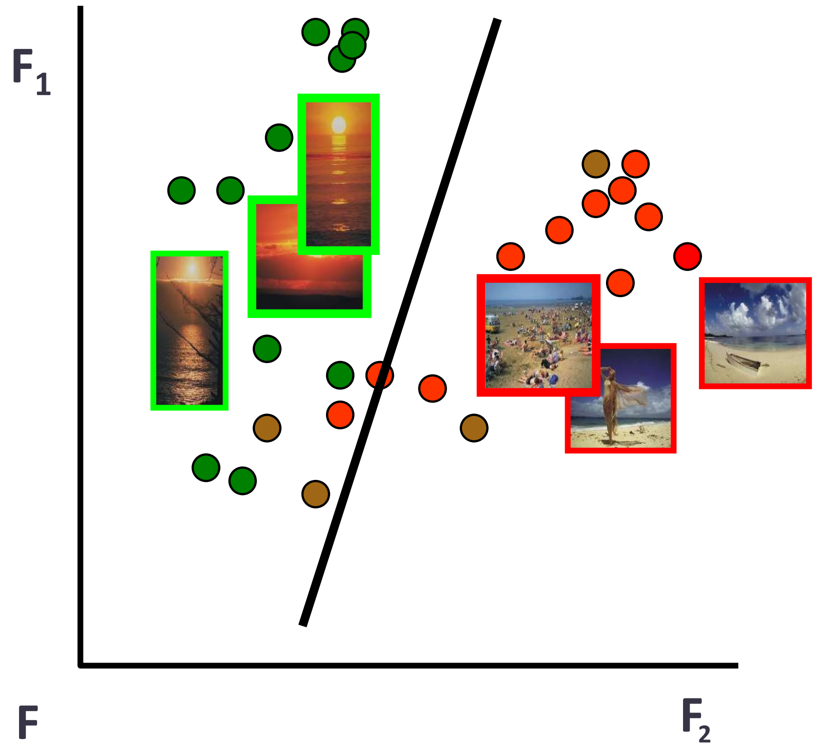
Relevance feedback



Try to find boundary in feature space best separating positive from negative examples

 **Measure of class membership probability**

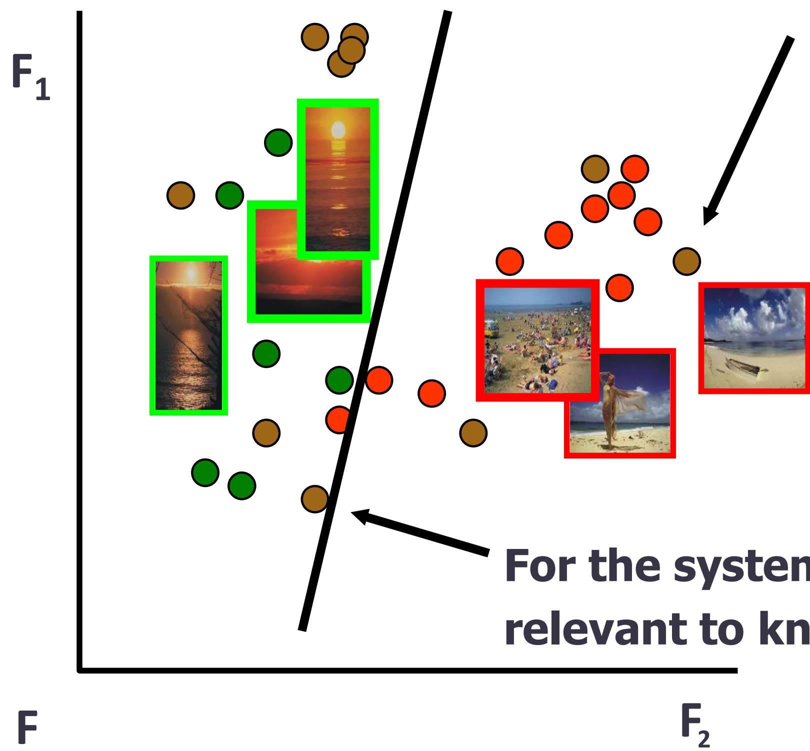
Relevance feedback



In the next iteration I will have more samples hence a better estimate of the boundary

This process is usually known as relevance feedback

Active Learning

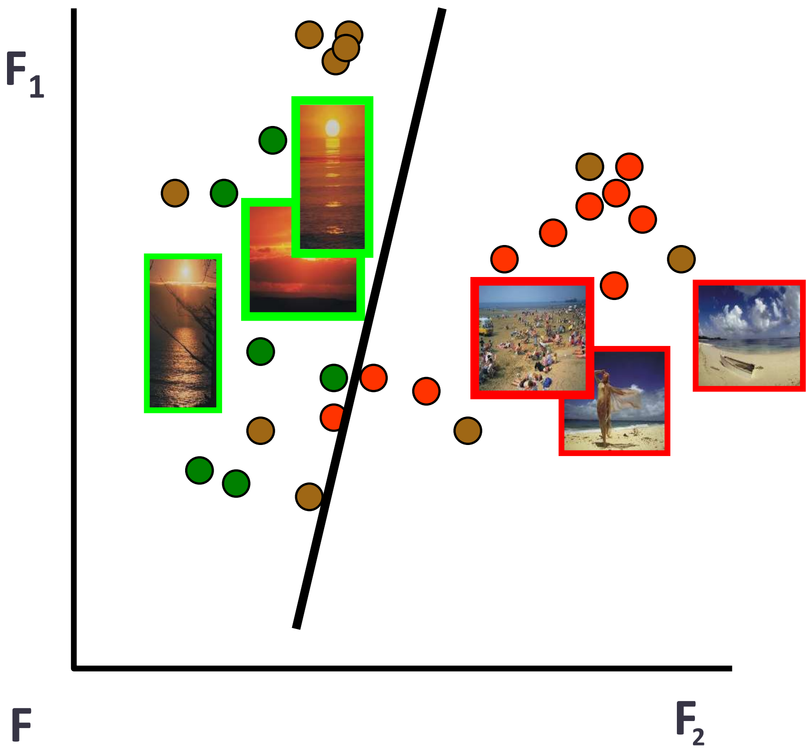


The system can safely assume this sample is also negative

In active learning the system decides which elements to show for feedback and which not.

For the system it is relevant to know this label

Before deep learning

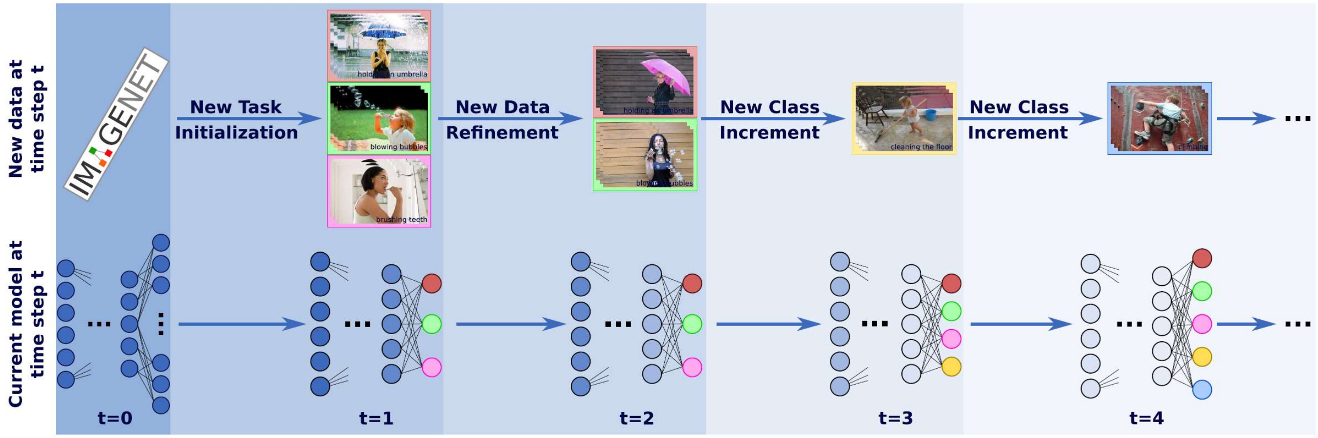


Linear or non-linear Support Vector Machine based on all current positive and negative examples

Use output of the classifier to compute probabilities and either give best output (RF) or actively select elements for labeling (AL)

With deep learning

- Two approaches
  - Apply SVM to latent spaces of the model
  - Fine-tuning deep learned models

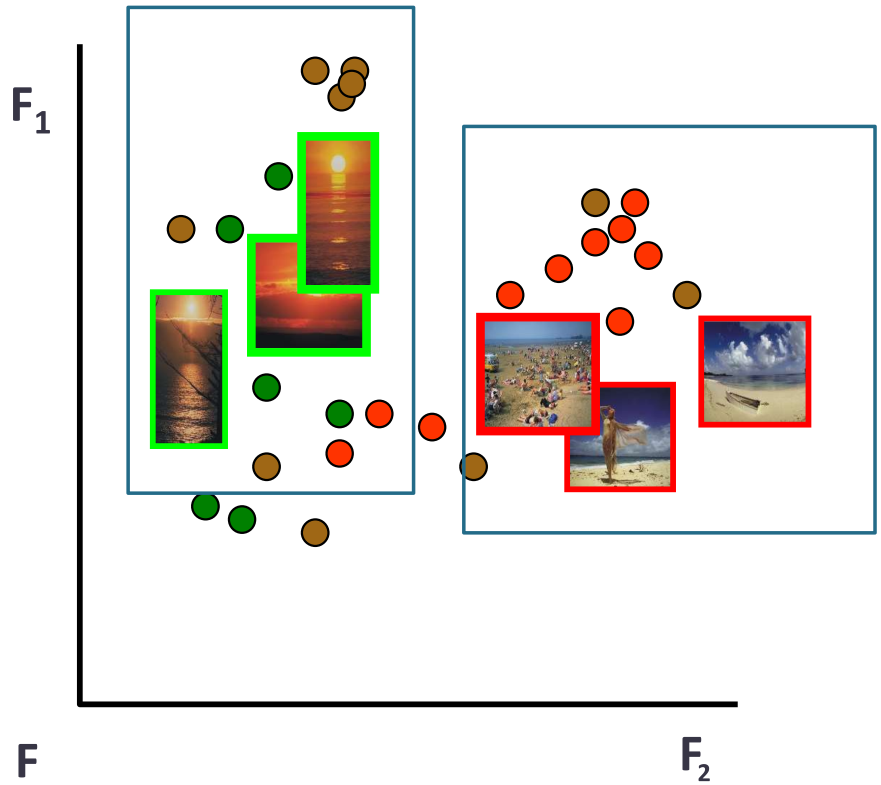


Main challenge is making this real time

[Kading2016]

Example Machine Teaching

First learn to recognize the easy sunsets



Then the more difficult beaches

Very much fits the model just presented

Relevance feedback or Active Learning?

- Active learning – suggest items the model is *least sure* about
  - Those nearest to the decision boundary
- Adds the most useful information to the model at the cost of lowest number of labels
  - After training the model, it will be used to auto-label the rest of the data
  - Saving the user's time in the long run
- But the user doesn't see crisply-relevant items
  - So, the value of the model may be in doubt
- And for some items, it's impossible to decide even for the user
- So, relevance feedback is still what you see most

Explanatory Interactive Learning

Algorithm 1 CAIPI takes as input a set of labelled examples $\mathcal{L}$, a set of unlabelled instances $\mathcal{U}$, and iteration budget T.

$f \leftarrow \text{FIT}(\mathcal{L})$
repeat
  $x \leftarrow \text{SELECTQUERY}(f, \mathcal{U})$
  $\hat{y} \leftarrow f(x)$
  $\hat{z} \leftarrow \text{EXPLAIN}(f, x, \hat{y})$
  Present x, $\hat{y}$, and $\hat{z}$ to the user
  Obtain y and explanation correction $\mathcal{C}$
  $\{(\bar{x}_i, \bar{y}_i)\}_{i=1}^{c} \leftarrow \text{TOCOUNTEREXAMPLES}(\mathcal{C})$
  $\mathcal{L} \leftarrow \mathcal{L} \cup \{(x, y)\} \cup \{(\bar{x}_i, \bar{y}_i)\}_{i=1}^{c}$
  $\mathcal{U} \leftarrow \mathcal{U} \setminus (\{x\} \cup \{\bar{x}_i\}_{i=1}^{c})$
  $f \leftarrow \text{FIT}(\mathcal{L})$
until budget T is exhausted or f is good enough
return f

During interactions three cases can occur:
(1) Right for the right reasons: The prediction and the explanation are both correct. No feedback is requested.
(2) Wrong for the wrong reasons: The prediction is wrong. As in active learning, we ask the user to provide the correct label. The explanation is also necessarily wrong, but we currently do not require the user to act on it.
(3) Right for the wrong reasons: The prediction is correct, but the explanation is wrong. We ask the user to provide an explanation correction C.

The user changes attributes that are wrong with random / other values thus creating counterexamples, system should learn to ignore those

[Teso2019]

Difference Active Learning and IML

- Both focus on selecting new points for labeling by the user
- Active learning
  - Selection is driven by the model
- Interactive Machine Learning
  - Selection is driven by the user

Interactive Machine Learning: Main features

- Humans in the ML loop
  - assigned to tasks at which they are more efficient than machines.
- Humans assuming different roles
  - which can be domain experts, non-expert users, data scientists, etc.
- Incremental methodology
  - in which the model is updated iteratively and incrementally.
- The importance of the user interface
  - which influences how learning takes place and conditions learning outcomes.

Focus on much more diverse tasks than only classification

Tasks in Interactive Machine Learning

- Visual cluster analysis
  - Exploratory data analysis
  - Comparative clustering analysis
  - Bi-cluster analysis (rows and columns in a matrix)
- Interactive dimensionality reduction
  - Subspace analysis
  - High-dimensional data exploration
  - Progressive dimensionality reduction
- Interactive model analysis
  - Model understanding
  - Model diagnosis

Tasks in Interactive Machine Learning

- Interactive classification
  - Interactive labeling
  - Interactive feature engineering
  - Parameter space analysis
- Interactive regression
  - Interactive correlation analysis
  - Interactive numerical prediction
- Interactive information retrieval
  - Structured information retrieval
  - Unstructured information retrieval

Tasks in Interactive Machine Learning

- Visual pattern mining
  - Exploratory event analysis
  - Mobility pattern analysis
- Visual topic analysis
  - Flat topic analysis
  - Hierarchical topic analysis
  - Topic evolution analysis
- Interactive anomaly detection
  - Anomalous points detection
  - Anomalous sequences detection

On-line learning

Online learning represents a family of machine learning methods, where a learner attempts to tackle some predictive (or any type of decision-making) task by learning from a sequence of data instances one by one at each time. The goal of online learning is to maximize the accuracy/correctness for the sequence of predictions/ decisions made by the online learner given the knowledge of correct answers to previous prediction/ learning tasks and possibly additional information.

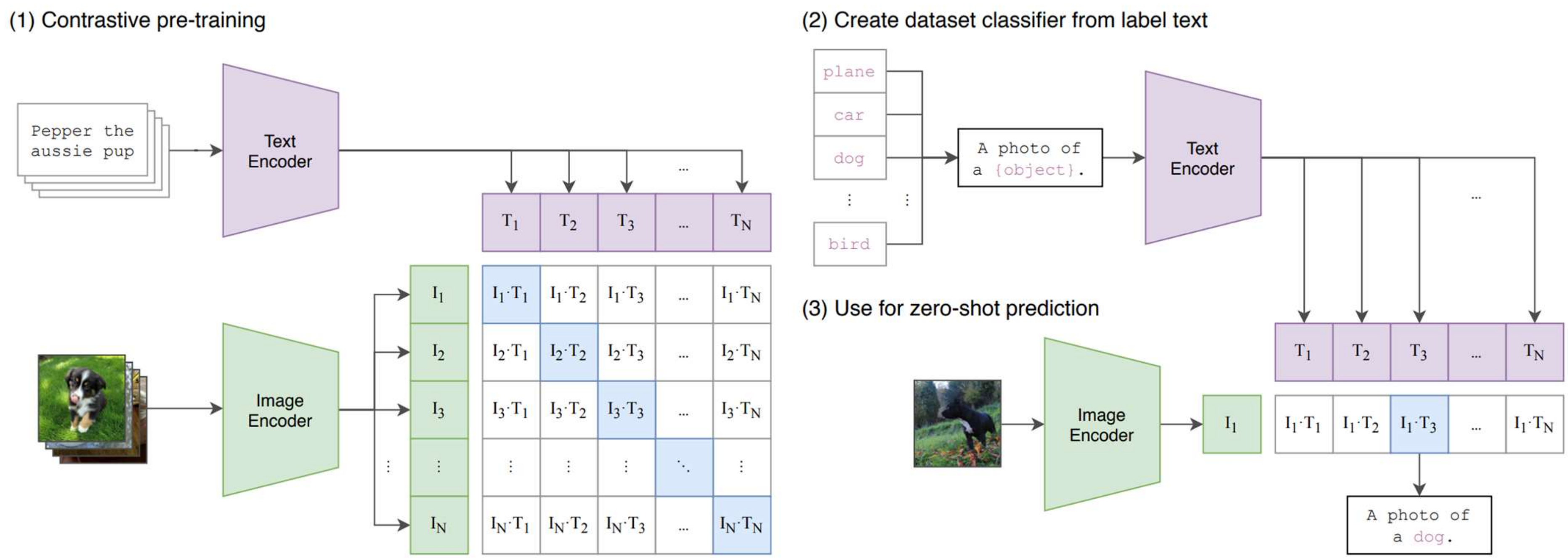
[Hoi2021]

On-line learning: overview

Online Learning			
Statistical Learning Theory	Convex Optimization Theory		Game Theory
Online Learning with Full Feedback		Online Learning with Partial Feedback (Bandits)	
Online Supervised Learning		Stochastic Bandit	Adversarial Bandit
First-order Online Learning	Online Learning with Regularization	Stochastic Multi-armed Bandit	Adversarial Multi-armed Bandit
Second-order Online Learning	Online Learning with Kernels	Bayesian Bandit	Combinatorial Bandit
Prediction with Expert Advice	Online to Batch Conversion	Stochastic Contextual Bandit	Adversarial Contextual Bandit
Applied Online Learning		Online Active Learning	Online Semi-supervised Learning
Cost-Sensitive Online Learning	Online Collaborative Filtering	Selective Sampling	Online Manifold Regularization
Online Multi-task Learning	Online Learning to Rank	Active Learning with Expert Advice	Transductive Online Learning
Online Multi-view Learning	Distributed Online Learning	Online Unsupervised Learning (no feedback)	
Online Transfer Learning	Online Learning with Neural Networks	Online Clustering	Online Density Estimation
Online Metric Learning	Online Portfolio Selection	Online Dimension Reduction	Online Anomaly Detection

Many methods for different purposes, good inspiration for IML

How about Zero shot learning?



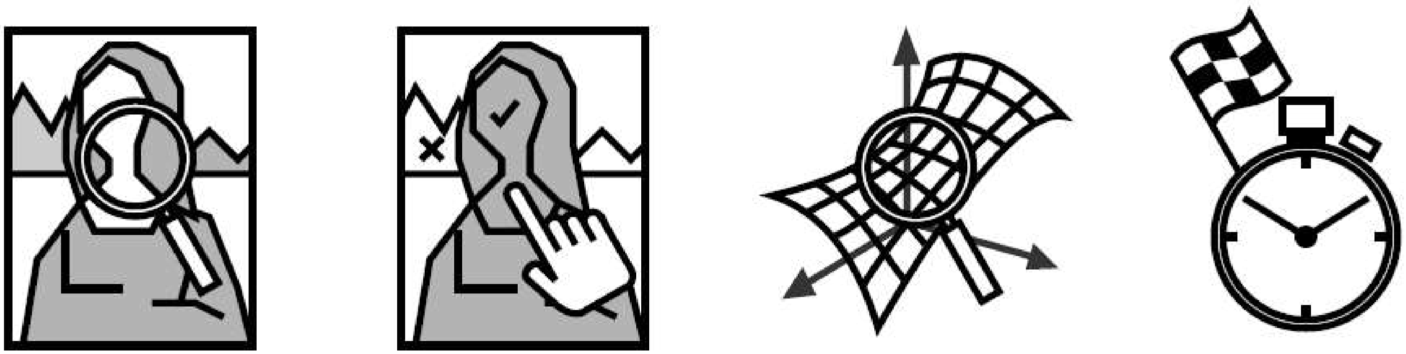
Uses the objective learned context, but doesn't adapt to the user

[Radford2021]

User Interface Design

[DUDLEY 2018]

Distinct Interface Elements



Sample Review

Visualization of output sample(s) to assess how well desired concept operates at the instance level.

Feedback Assignment

Assignment or correction of labels and/or creation of new samples to improve match with desired concept.

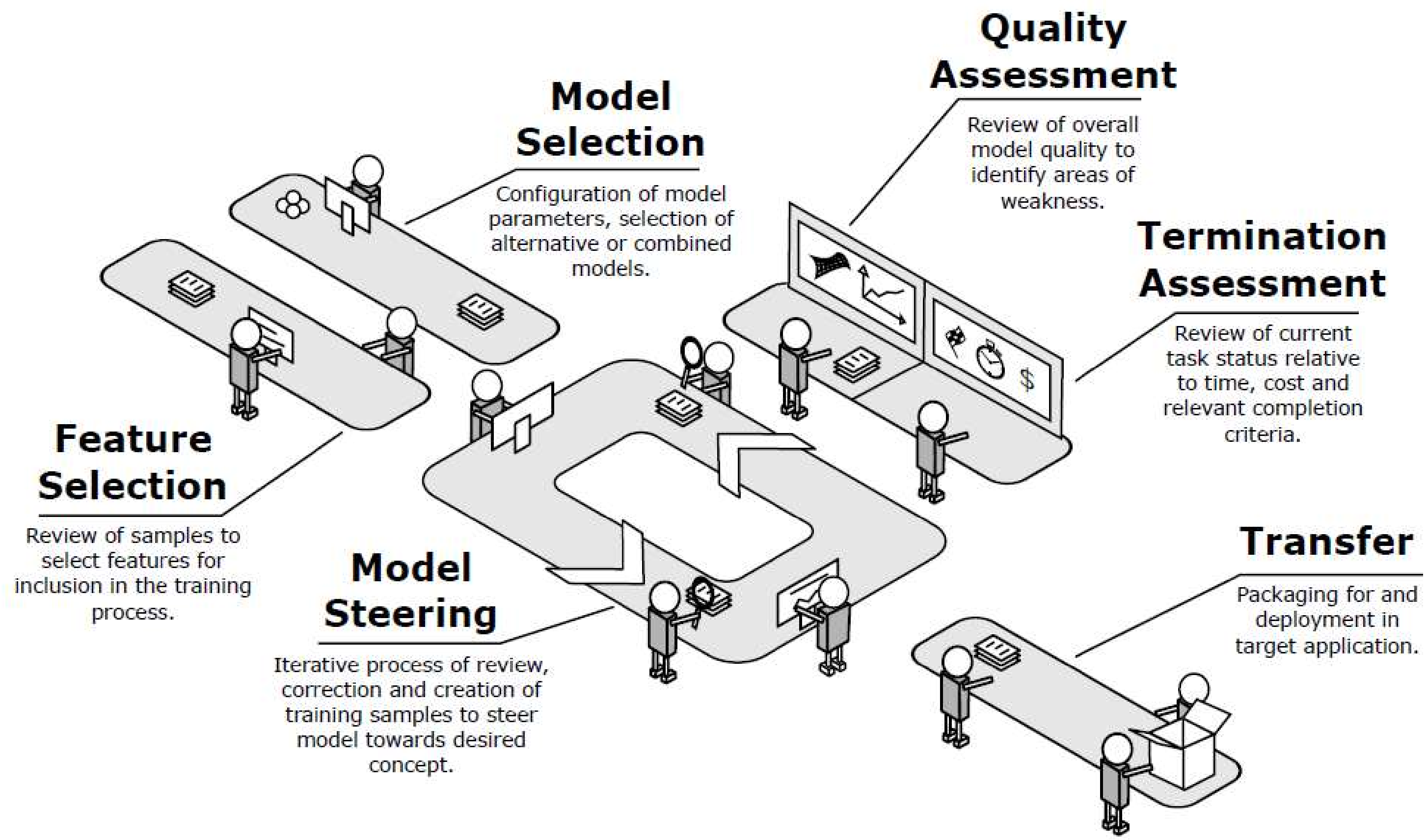
Model Inspection

View of overall model quality and converage to evaluate how well concept is captured.

Task Overview

View of current task status contexualized by coverage of training data and improvement potential relative to cost.

Distinct User Activities



Six key emergent solutions

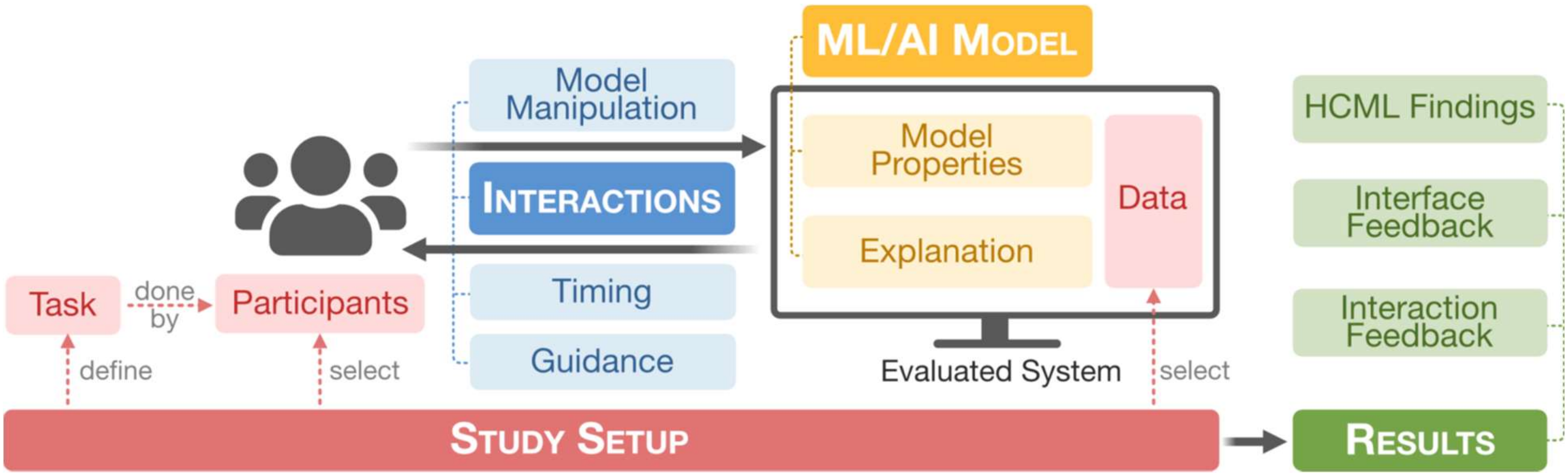
- Make task goals and constraints explicit
- Support user understanding of model uncertainty and confidence
- Capture intent rather than input
- Provide effective data representations
- Exploit interactivity and promote rich interactions
- Engage the user

Evaluation of Interactive Learning

Evaluation

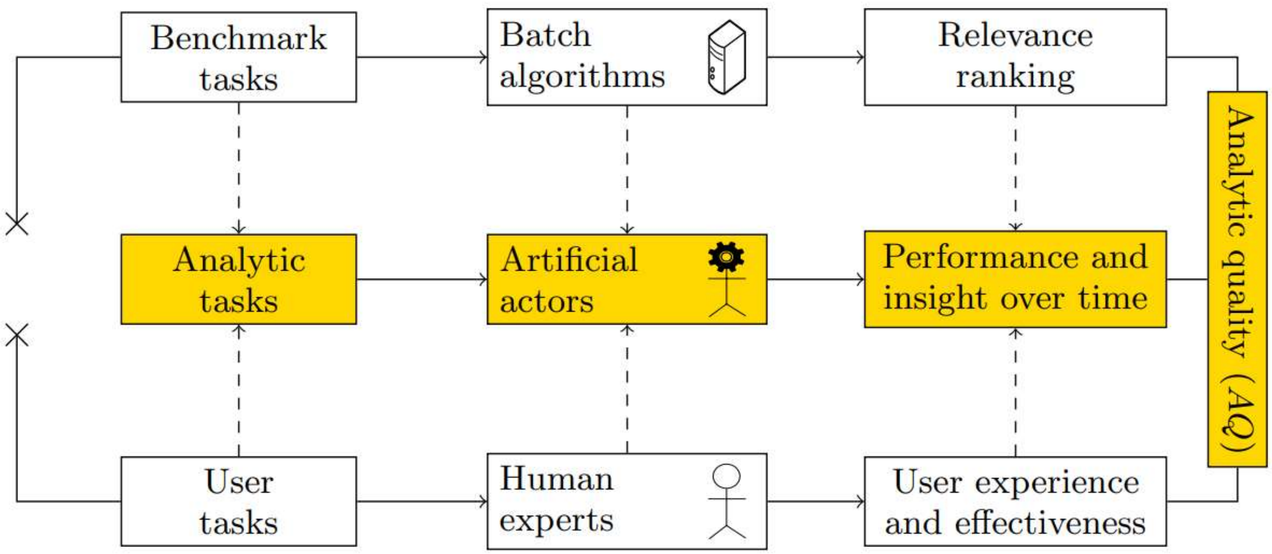
- What we need to do
  - Evaluate the quality of the model
  - Evaluate the user experience
  - Evaluate effort from both user and system

Many methods for evaluation



[Sperlle2021]

Analytic Quality

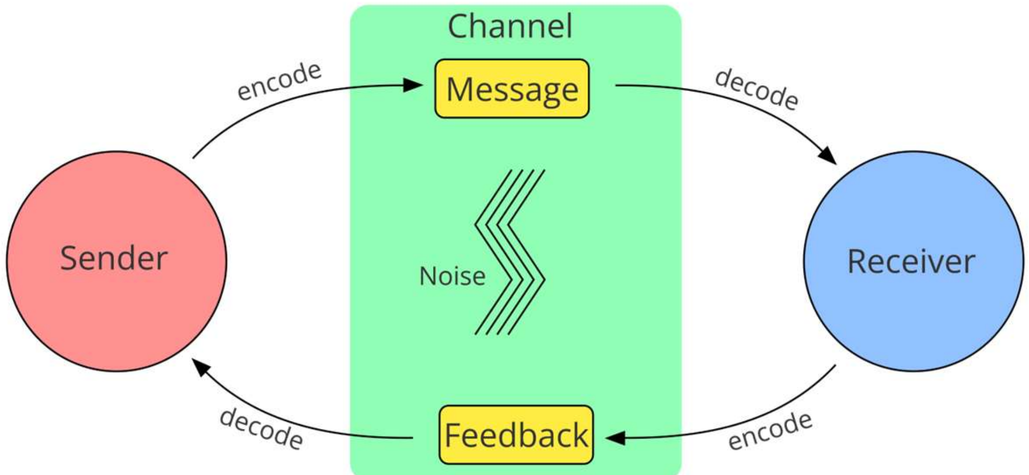


[Zahalka2015]

Visualization

Bringing Machine Learning and Users Together

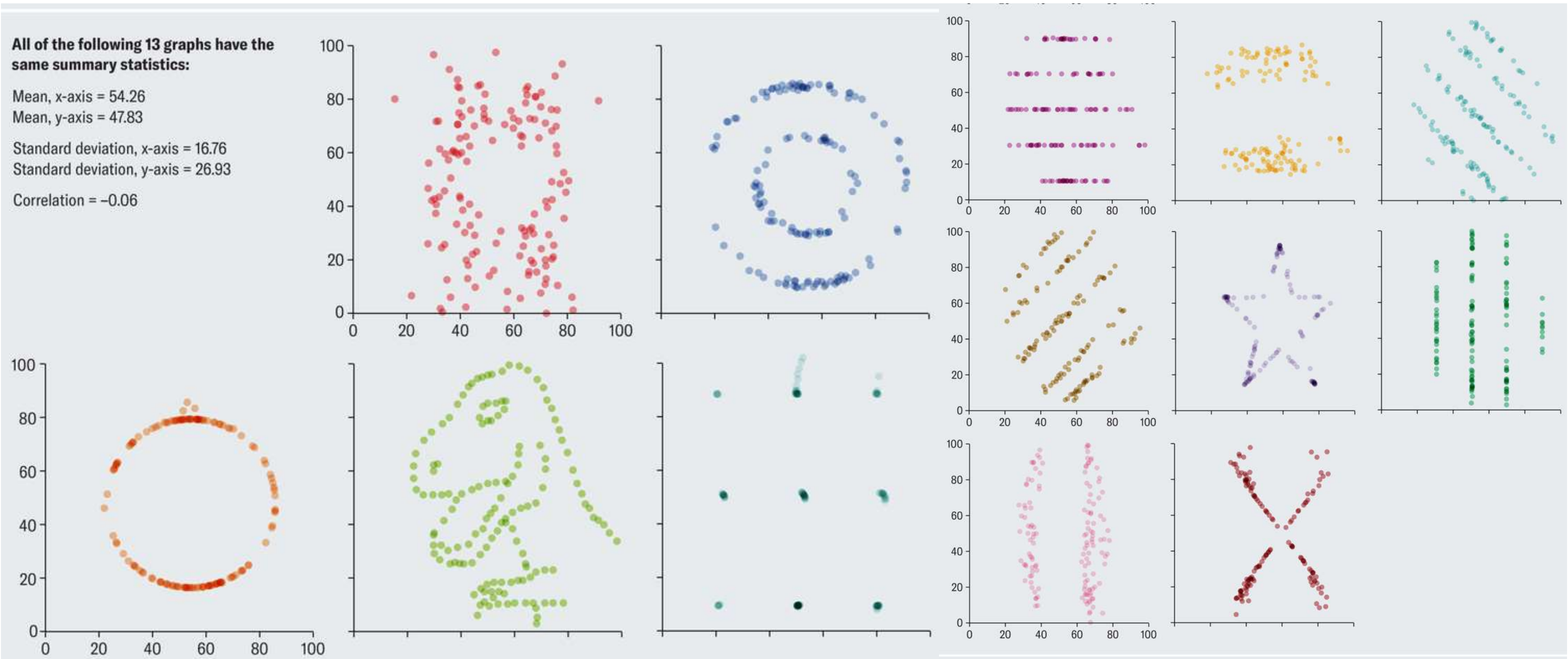
- Requires communication channels between the computer and the human



The brain uses >50% for visual processing so visualization is a natural way to communicate, so why use the very limited channel of prompting?

Figure from https://en.wikipedia.org/wiki/Communication_theory

Why visualization?

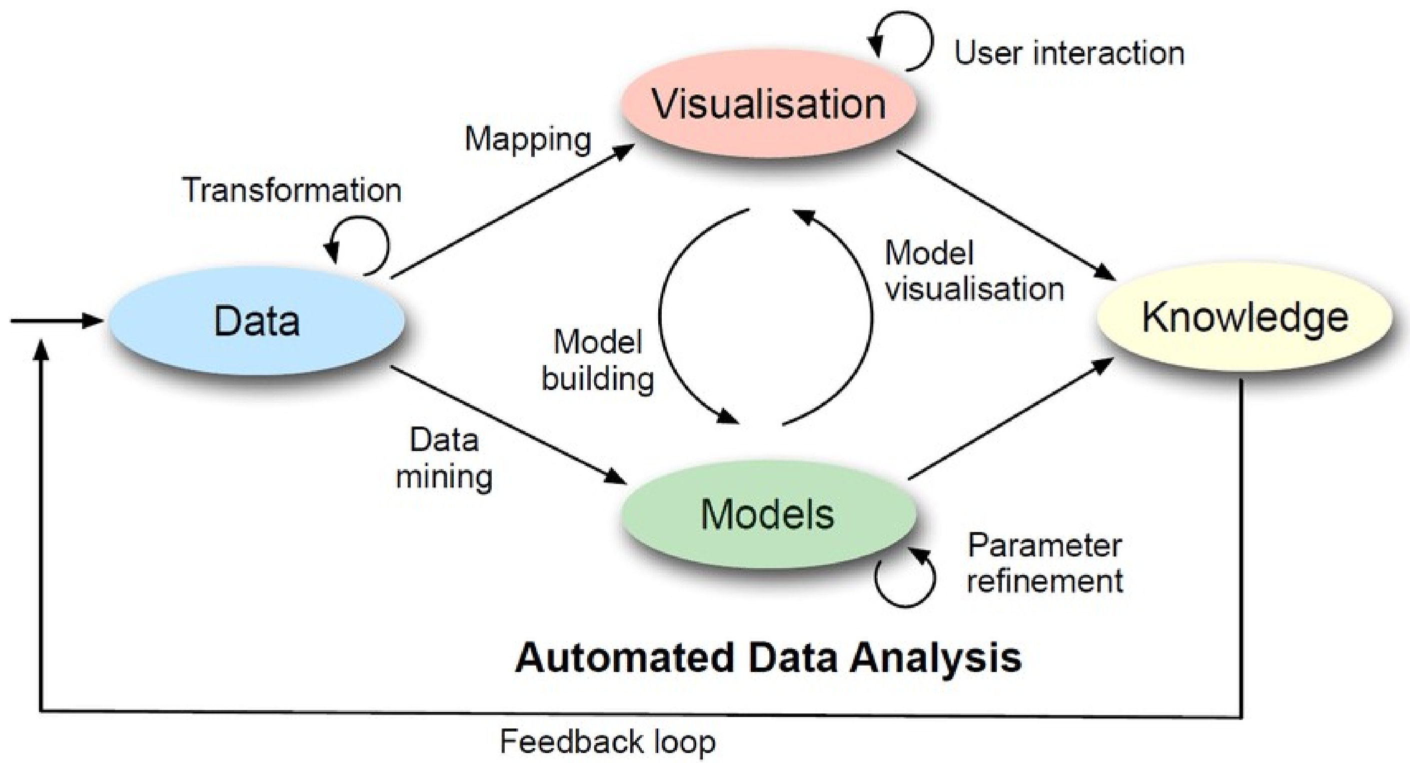


The best fitting line for all of them would be the same!

https://www.scientificamerican.com/article/what-this-graph-of-a-dinosaur-can-teach-us-about-doing-better-science/

Visual Analytics

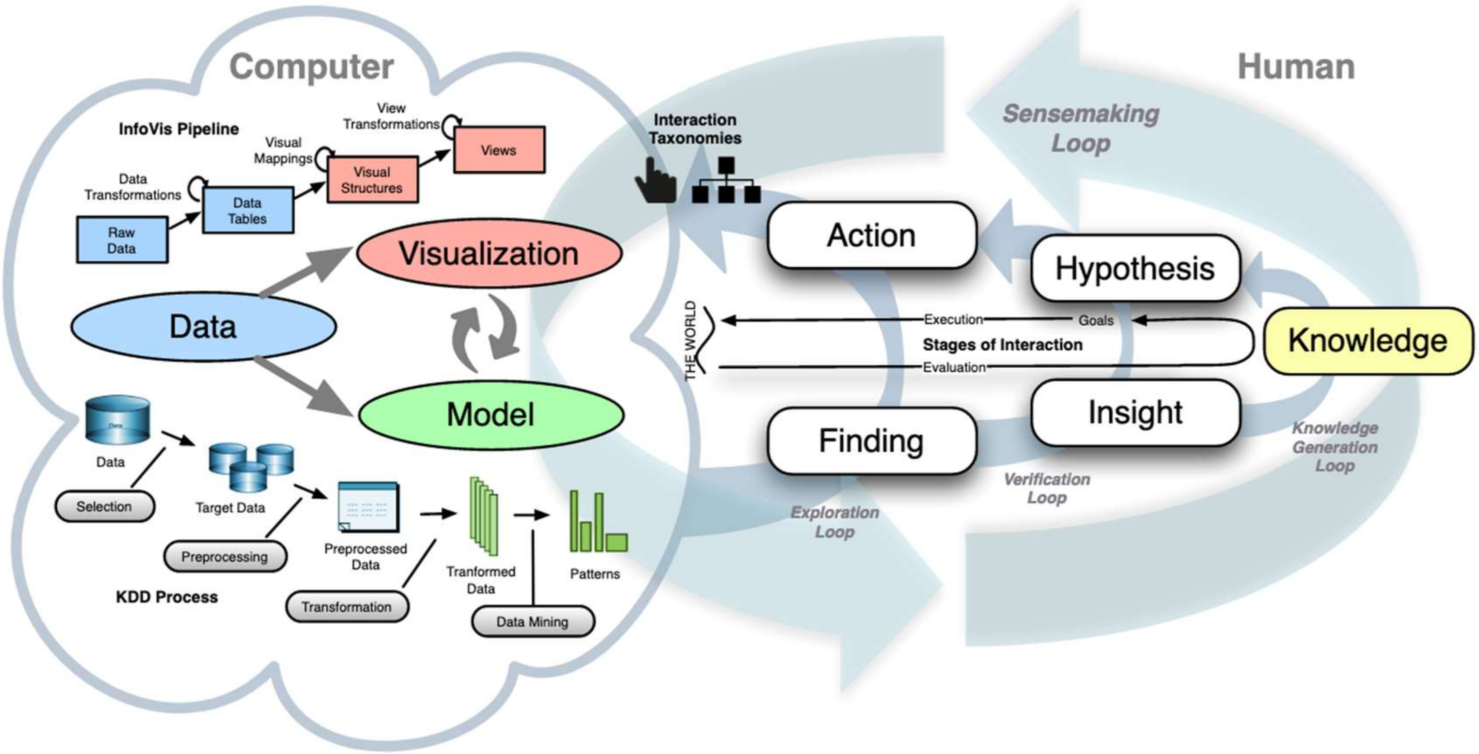
The basic visual analytics model



Bringing together information visualization and data mining

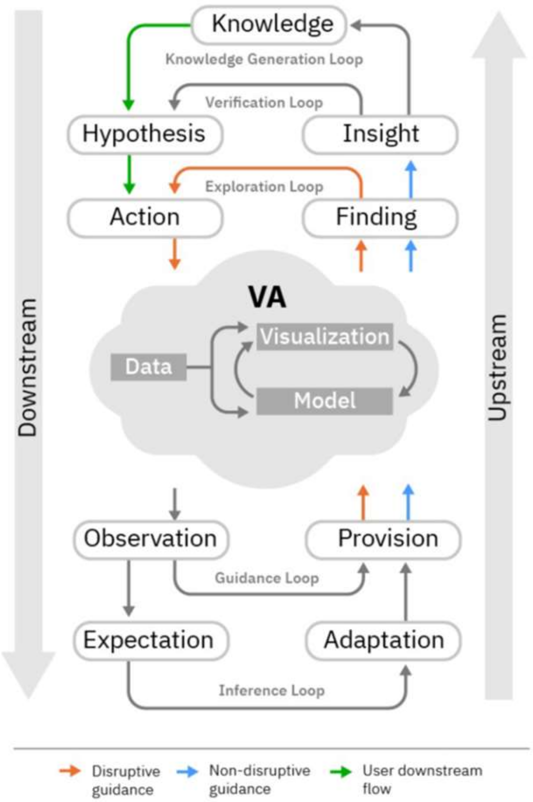
[Keim2008]

VA pipeline + human: Fully expanded



A comprehensive VA model integrating several, previously isolated, models [Sacha14]

A user guidance adaption of the model



More focus on the user side, hence closer to IML

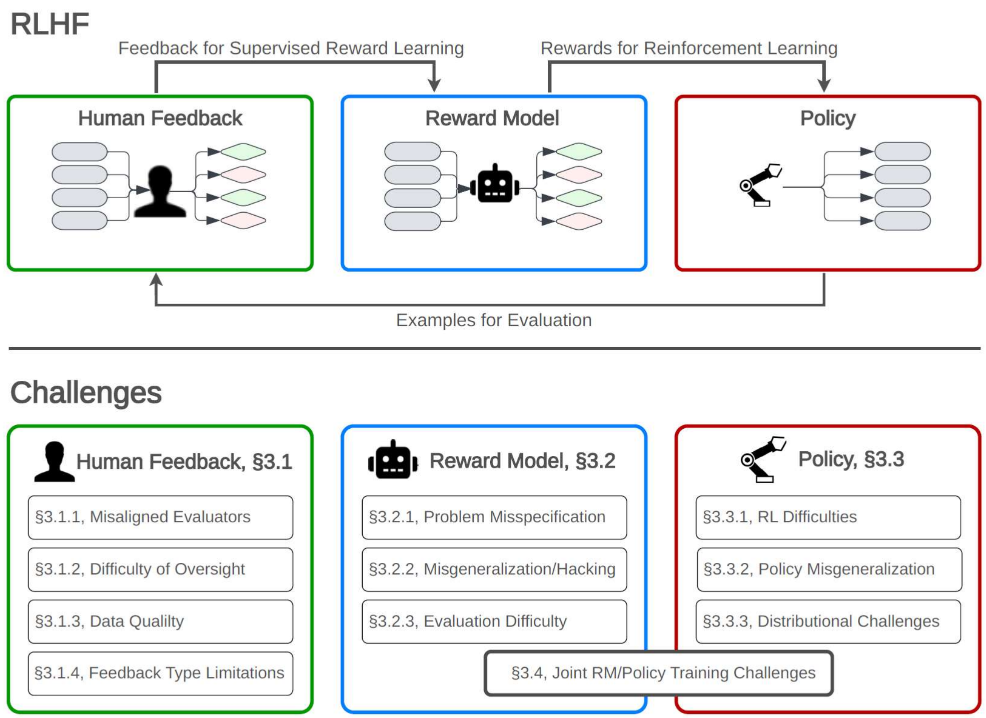
[Perez-Messina2022]

VA Model: Modelling technique categories

- Modify parameters (MP)
  - The user directly manipulates the model parameters through the visualization
  - The more populous category across all techniques
  - Pros: easier to implement, exact meaning
  - Cons: requires stats/machine learning knowledge from the user, non-intuitive
- Define analytical expectations (DAE)
  - The user interacts within the domain of expertise (using domain knowledge), the model behaves semantically: translating between the user's language and the ML/stats language
  - Fewer approaches exist
  - Pros: meaningful and intuitive to the user, no or little knowledge of stats/ML required
  - Cons: difficult to implement, knowledge gap between the developer and the user

[Endert17]

Reinforcement learning seems good model for DAE



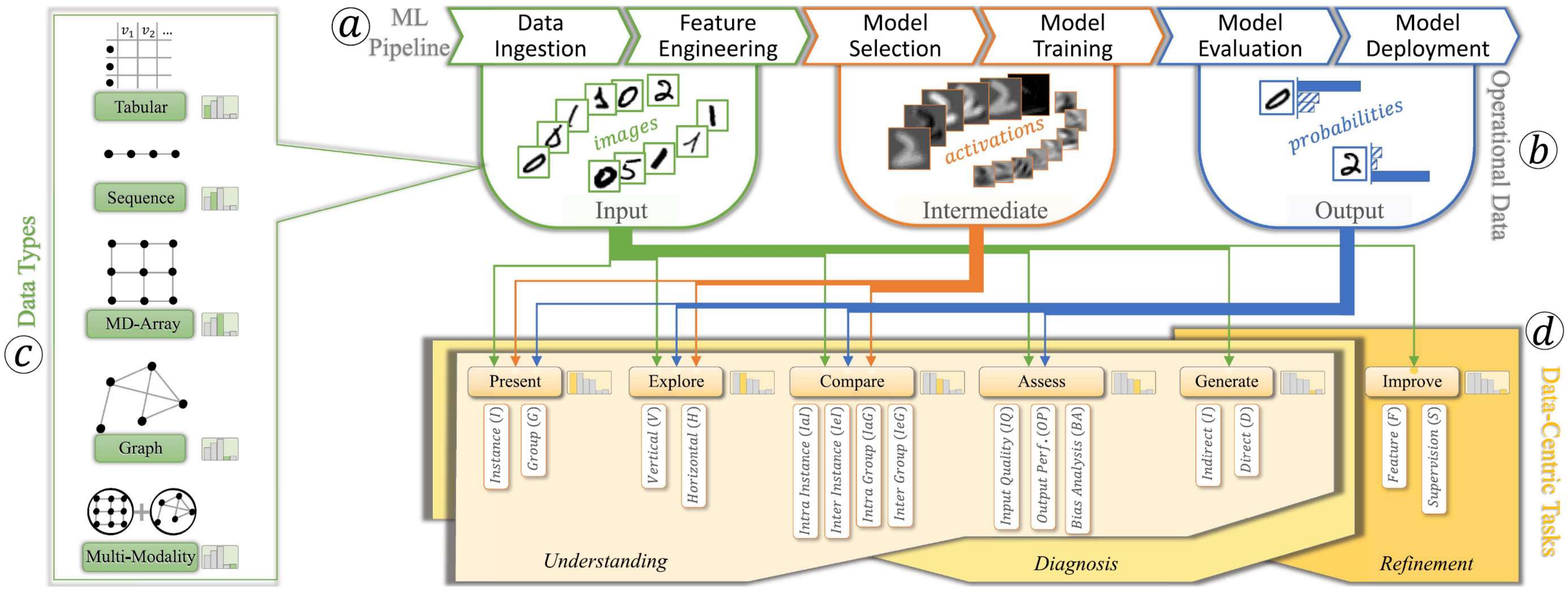
But has its challenges, more on RLHF in subsequent lecture

[Casper2023]

Visual Analytics: a Data Perspective

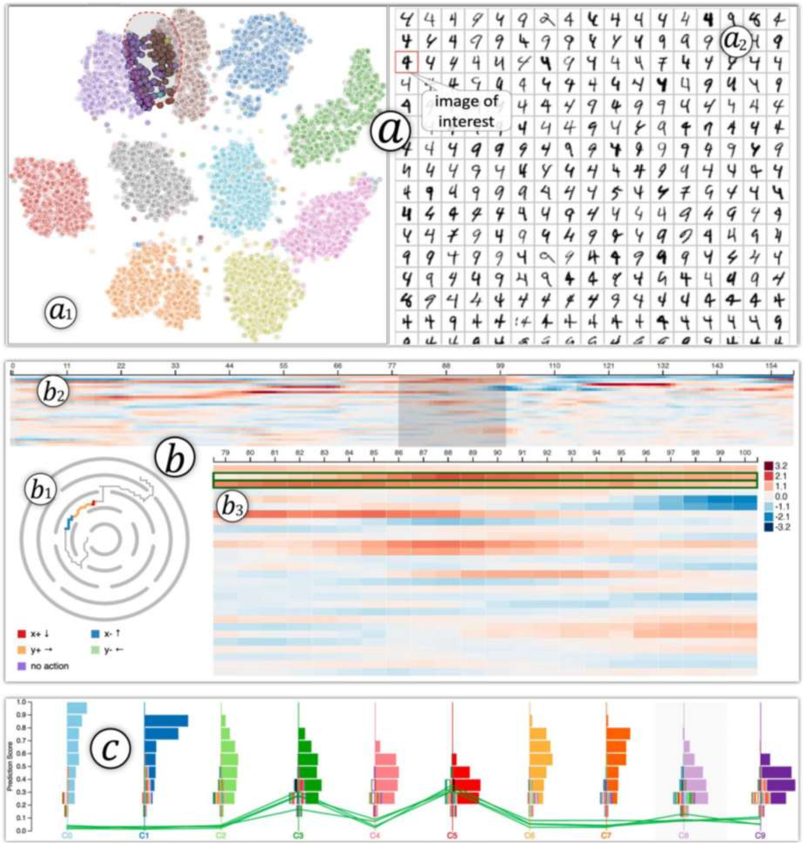
[WANG2024]

Overview



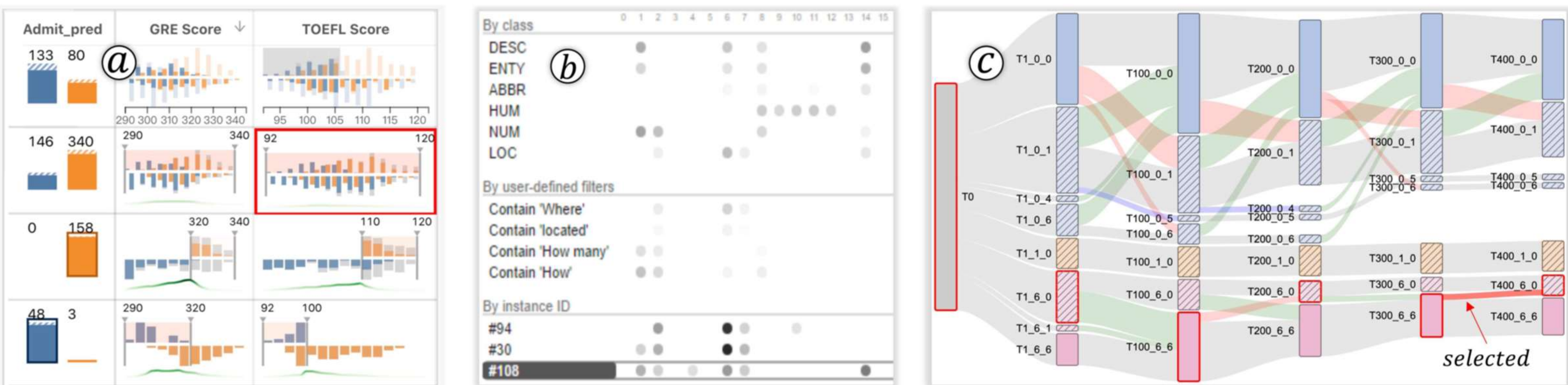
Data Types
Tabular
Sequence
MD-Array
Graph
Multi-Modality
ML Pipeline
Data Ingestion
Feature Engineering
Model Selection
Model Training
Model Evaluation
Model Deployment
images
Input
activations
Intermediate
probabilities
Output
Operational Data
Present
Explore
Compare
Assess
Generate
Improve
Instance (I)
Group (G)
Vertical (V)
Horizontal (H)
Intra Instance (IaI)
Inter Instance (IeI)
Intra Group (IaG)
Inter Group (IeG)
Input Quality (IQ)
Output Perf. (OP)
Bias Analysis (BA)
Indirect (I)
Direct (D)
Feature (F)
Supervision (S)
Understanding
Diagnosis
Refinement
Data-Centric Tasks

Instance Level Presentation



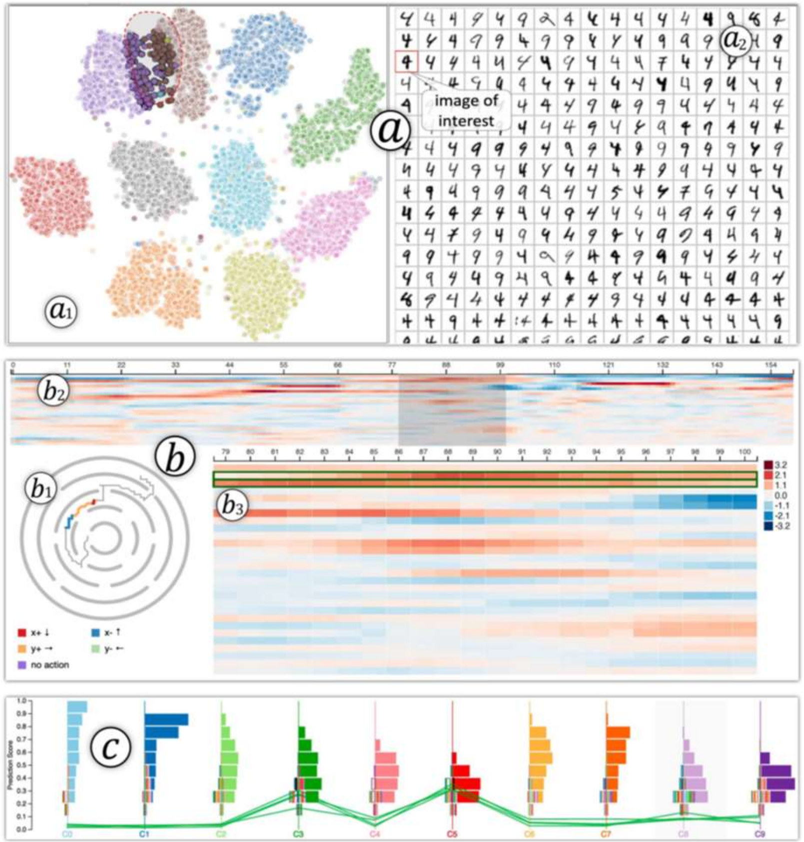
(a) Input images are presented through a scatterplot, one point for one image.
(b) Intermediate hidden states are externalized through a heatmap, each row is an instance (principle component) and each column is a time step.
(c) Output probabilities presented through a parallel coordinates plot, one polyline per instance

Group Level Presentation



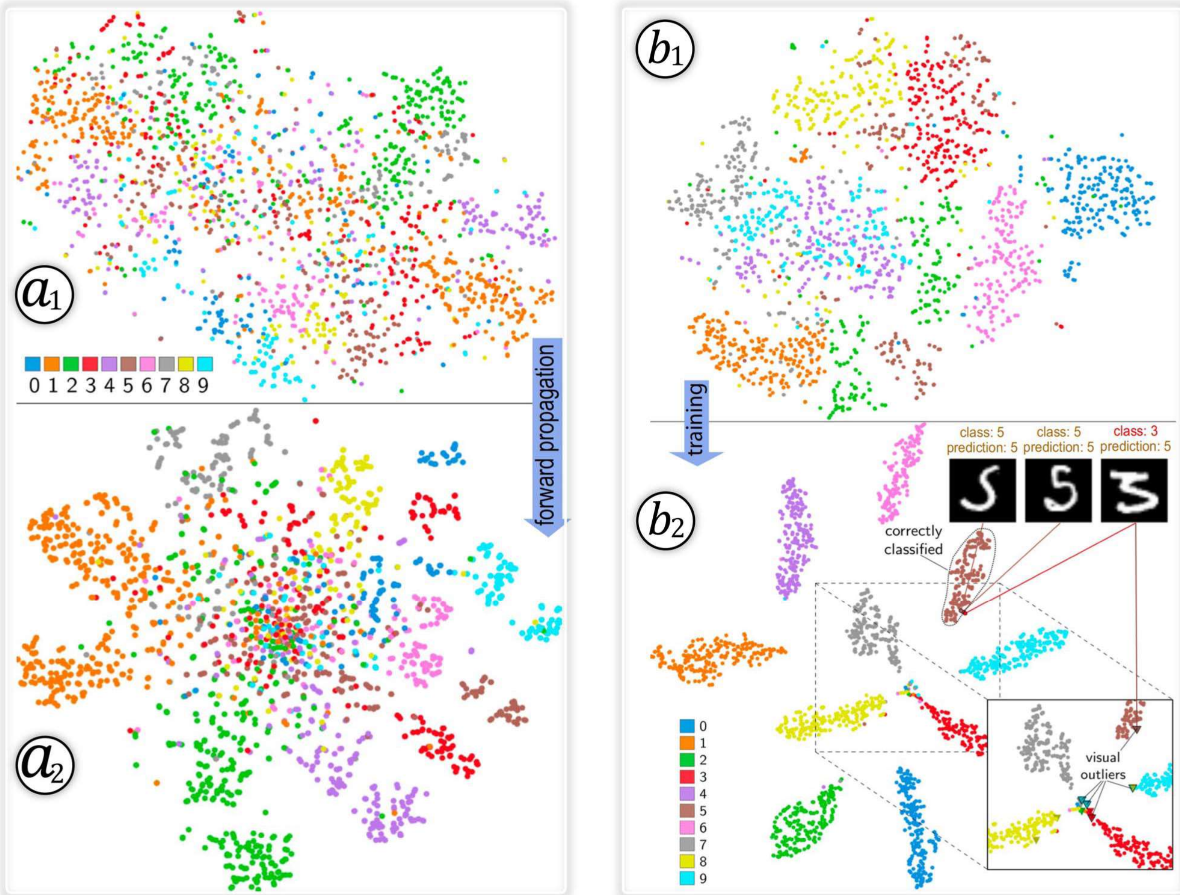
(a) The tabular input data in DECE are divided into subgroups and presented as histograms.
(b) The intermediate DNN activations from subgroups of instances are aggregated in and presented as circles, whose color denotes the active level.
(c) A Sankey-diagram based temporal confusion matrix is used to present the output prediction over the training of a tree-boosting model.

Vertical Exploration



This is also an example of vertical exploration

Horizontal Exploration

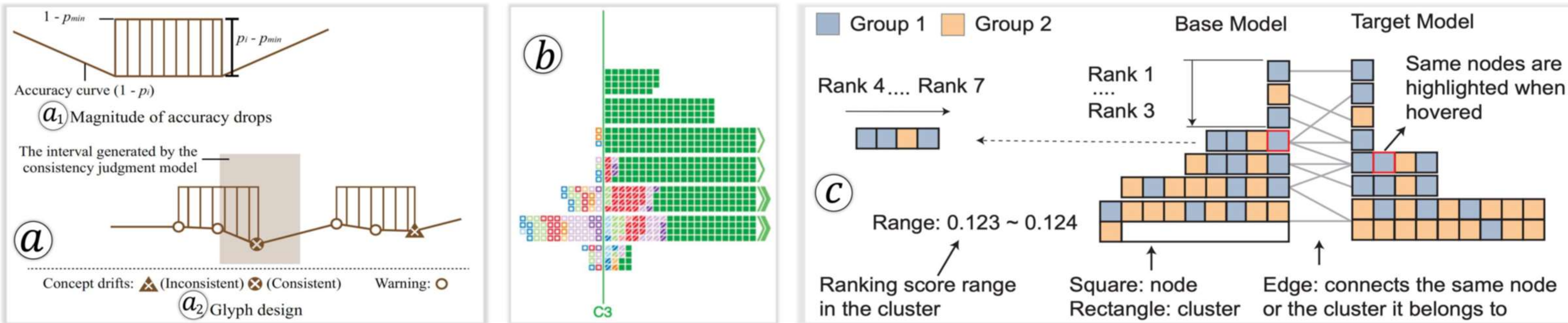


Explore data across different DNN layers (a1, a2) or training iterations (b1, b2).

Compare Data

(a) Intra-instance: SCANViz compares the reconstructions of the same image.
(b) Inter-instance: AEVis compares the datapaths of three images to diagnose adversarial attacks.
(c) Intra-group: EmbeddingVis compares the embeddings for the same group of instances from different models.
(d) Inter-group: FairVis compares model performance across instance groups.

Assess Data

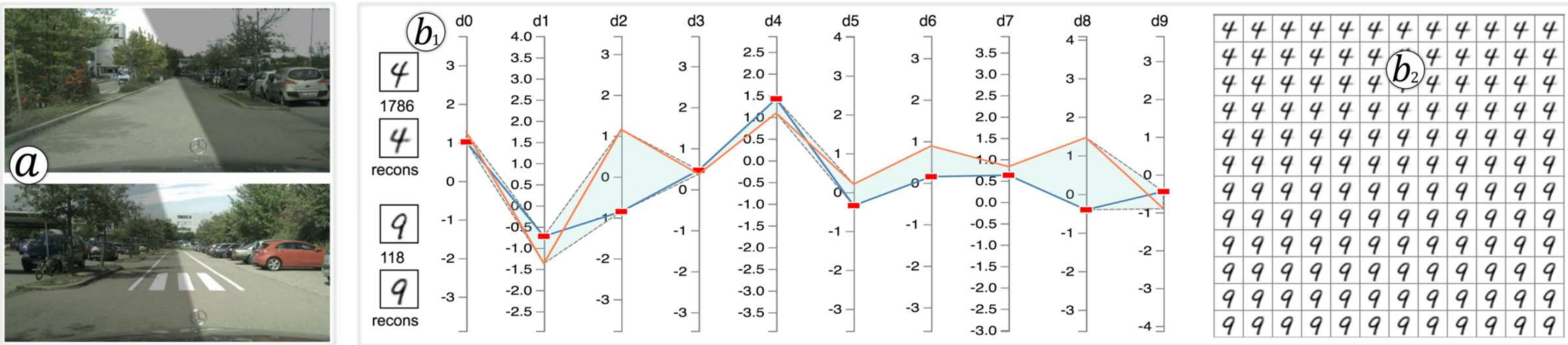


(a) Glyphs designed to identify concept drift.

(b) Each square represents one instance and its vertical position shows the class probability. The square glyphs and their position also encode the prediction correctness.

(c) Graph nodes (in orange and blue) are clustered by their ranking score and nodes of the same cluster are presented in a rectangle for similar exposure. Rankings from two models can also be compared.

Generate Data

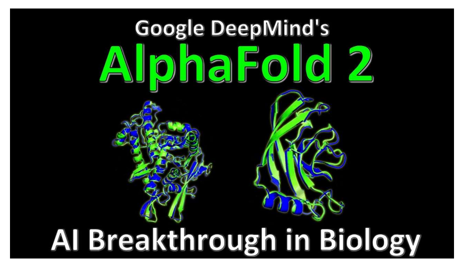


(a) Images are augmented by adding artificially generated shadows. DeepVID generates images between the to-be-interpreted digit ‘4’ and ‘9’ (b2) by interpolating their latent vectors (b1).

Artificial General Intelligence

Artificial General Intelligence

Performance (rows) x **Generality** (columns)	**Narrow** *clearly scoped task or set of tasks*	**General** *wide range of non-physical tasks, including metacognitive abilities like learning new skills*
Level 0: No AI	**Narrow Non-AI** calculator software; compiler	**General Non-AI** human-in-the-loop computing, e.g., Amazon Mechanical Turk
Level 1: Emerging *equal to or somewhat better than an unskilled human*	**Emerging Narrow AI** GOFAI[4]; simple rule-based systems, e.g., SHRDLU (Winograd, 1971)	**Emerging AGI** ChatGPT (OpenAI, 2023), Bard (Anil et al., 2023), Llama 2 (Touvron et al., 2023)
Level 2: Competent *at least 50th percentile of skilled adults*	**Competent Narrow AI** toxicity detectors such as Jigsaw (Das et al., 2022); Smart Speakers such as Siri (Apple), Alexa (Amazon), or Google Assistant (Google); VQA systems such as PaLI (Chen et al., 2023); Watson (IBM); SOTA LLMs for a subset of tasks (e.g., short essay writing, simple coding)	**Competent AGI** not yet achieved
Level 3: Expert *at least 90th percentile of skilled adults*	**Expert Narrow AI** spelling & grammar checkers such as Grammarly (Grammarly, 2023); generative image models such as Imagen (Saharia et al., 2022) or Dall-E 2 (Ramesh et al., 2022)	**Expert AGI** not yet achieved
Level 4: Virtuoso *at least 99th percentile of skilled adults*	**Virtuoso Narrow AI** Deep Blue (Campbell et al., 2002), AlphaGo (Silver et al., 2016, 2017)	**Virtuoso AGI** not yet achieved
Level 5: Superhuman *outperforms 100% of humans*	**Superhuman Narrow AI** AlphaFold (Jumper et al., 2021; Varadi et al., 2021), AlphaZero (Silver et al., 2018), StockFish (Stockfish, 2023)	**Artificial Superintelligence (ASI)** not yet achieved



M. Morris (Google DeepMind): Levels of AGI: Operationalizing Progress on the Path to AGI Arxiv 2023

Artificial General Intelligence

- GAIA, a benchmark for General AI Assistants with real-world questions that require a set of fundamental abilities such as reasoning, multi-modality handling, web browsing, and generally tool-use proficiency.

Level 1

Question: What was the actual enrollment count of the clinical trial on H. pylori in acne vulgaris patients from Jan-May 2018 as listed on the NIH website?
Ground truth: 90

Level 2

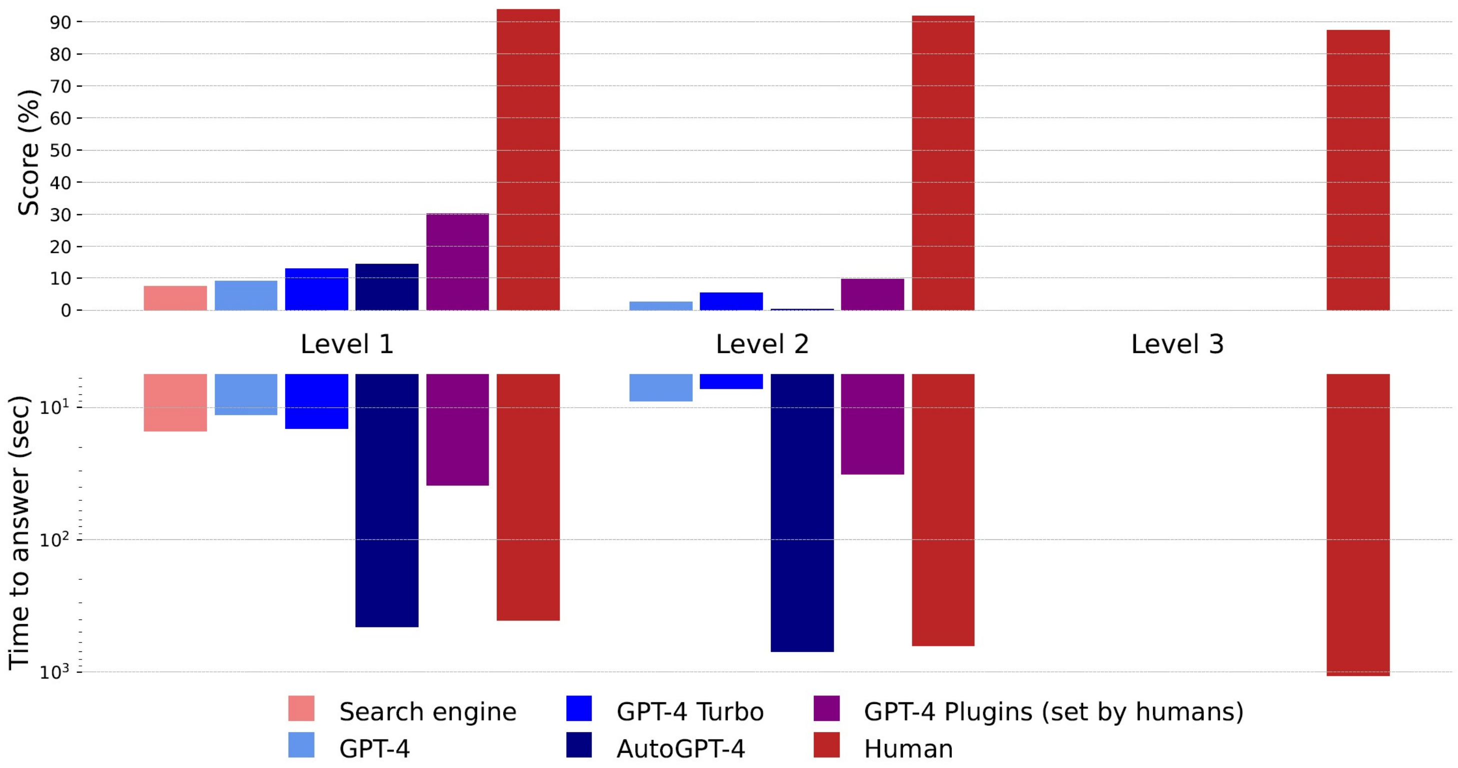
Question: If this whole pint is made up of ice cream, how many percent above or below the US federal standards for butterfat content is it when using the standards as reported by Wikipedia in 2020? Answer as + or - a number rounded to one decimal place.
Ground truth: +4.6

Level 3

Question: In NASA's Astronomy Picture of the Day on 2006 January 21, two astronauts are visible, with one appearing much smaller than the other. As of August 2023, out of the astronauts in the NASA Astronaut Group that the smaller astronaut was a member of, which one spent the least time in space, and how many minutes did he spend in space, rounded to the nearest minute? Exclude any astronauts who did not spend any time in space. Give the last name of the astronaut, separated from the number of minutes by a semicolon.
Ground truth: White; 5876

G. Mialon, C. Fourrier, C. Swift[3], T. Wolf, Y. LeCun, T. Scialom, ICLR 2024

A Comparison in performance



Artificial General Intelligence

- AI works well for specific well-defined applications
- On the GAIA benchmark
  - Humans perform better
  - Machines are faster
  - Multiple “agents/tool” needed to achieve the goal

Multi-Agent Systems

OpenAI Imagines Our AI Future

Stages of AI

Level 1	Chatbots, AI with conversational language
Level 2	Reasoners, human-level problem solving
Level 3	Agents, systems that can take actions
Level 4	Innovators, AI that can aid in invention
Level 5	Organizations, AI that can do the work of an organization

Source: Bloomberg reporting

Multi-Agent Systems

- Agents
  - Will play an increasingly important role in the future
  - Will be specialized to specific tasks
  - Can collaborate
- The role of interactive machine learning
  - At the individual agent level
  - At the whole system level
- Reinforcement learning
  - With appropriately set goals promised to be an interesting avenue to realize this
- Visual Analytics
  - New systems needed for this new context

Conclusions

Conclusions

- Usable and Useful AI
- Hybrid Intelligence needed
- Combination of ML, VIS and HCI needed
- LLMs and/or multi-agent systems
- Reinforcement learning a promising avenue
  - Explicit setting of goals
  - Multiple agents (including humans) aiming to reach the goals
- Many interesting research challenges

References

- [Casper2023] S. Casper et.al. Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. Transactions on Machine Learning Research (12/2023)
- [Dudley2018] Dudley JJ, Kristensson PO. A review of user interface design for interactive machine learning. ACM Transactions on Interactive Intelligent Systems, 2018
- [Endert2017] Endert, Alex, et al. "The state of the art in integrating machine learning into visual analytics." Computer Graphics Forum. Vol. 36. No. 8. 2017.
- [Hoi2021] Hoi, Steven CH, et al. "Online learning: A comprehensive survey." Neurocomputing 459 (2021): 249-289.
- [Kading2016] C. Kading, E. Rodner, A. Freytag, and J. Denzle. Fine-tuning Deep Neural Networks in Continuous Learning Scenarios. ACCV 2016
- [Keim08] D. Keim et al.: Visual Analytics: Definition, Process, and Challenges. In Information Visualization, Lecture Notes in Computer Science, vol. 4950, Springer, Berlin.
- [Mosqueria2022] E. Mosqueira-Rey et.al Human-in-the-loop machine learning: a state of the art. Artificial Intelligence Review (2023) 56:3005–3054
- [Perez-Messina22] Pérez-Messina, D. Ceneda, M. El-Assady, S. Miksch and F. Sperrle. A Typology of Guidance Tasks in Mixed-Initiative Visual Analytics Environment, Eurovis 2022
- [Radford21] Radford et.al. Learning Transferable Visual Models From Natural Language Supervision, ICML2021
- [Sacha14] D. Sacha et al.: Knowledge generation model for visual analytics. IEEE TVCG, 20 (12), pp. 1604 – 1613, December 2014.

References

- [Radford21] Radford et.al. Learning Transferable Visual Models From Natural Language Supervision, ICML2021
- [Sacha14] D. Sacha et al.: Knowledge generation model for visual analytics. IEEE TVCG, 20 (12), pp. 1604 – 1613, December 2014.
- [Sperlle2021] F. Sperrle, M. El-Assady, A Survey of Human-Centered Evaluations in Human-Centered Machine Learning. Guo, R. Borgo, D. Horng Chau, A. Endert, and D. Keim. EUROVIS 2021. Volume 40 (2021), Number 3. STAR – State of The Art Report
- [Teso2019] Teso, Stefano, and Kristian Kersting. "Explanatory interactive machine learning." Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. 2019.
- [Wang2024] Wang, Junpeng, Shixia Liu, and Wei Zhang. "Visual analytics for machine learning: A data perspective survey." IEEE Transactions on Visualization and Computer Graphics (2024).
- [Xu2019] Xu W Toward human-centered AI: a perspective from human–computer interaction. Interactions
- 26(4):42–46.
- [Zahalka2015] J Zahálka, S Rudinac, M Worring. Analytic quality: evaluation of performance and insight in multimedia collection analysis. Proceedings of the 23rd ACM international conference on Multimedia, 231-240