Hybrid intelligence (HI)

Interactive and semi-autonomous learning

Jan Zahálka

jan.zahalka@cvut.cz

Ω

Hybrid Intelligence (HI)

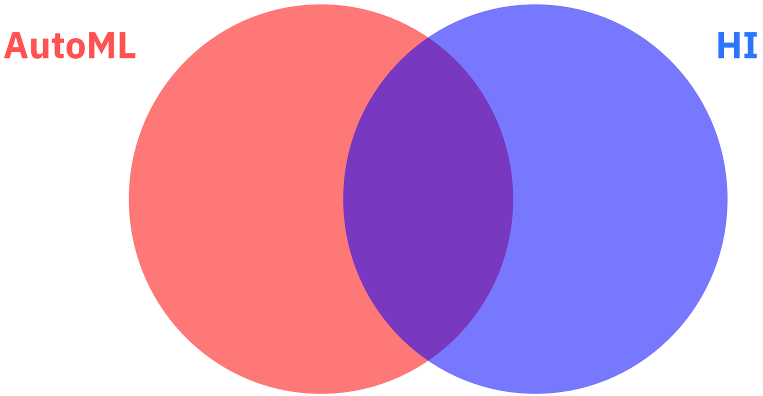
- AI that **interacts with a human** during the learning/task execution process and **learns from those interactions**
- HI = AI + human intelligence
- **Assists humans**, instead of **replacing** them
- This talk: my approach towards HI from the perspective of a person that does **machine learning** and has **multimedia analytics** experience
    - I am well aware that **AI/HI ≠ ML**

HI: Motivation

- Automated machine learning (AutoML) approaches already **assist** or **complete tasks autonomously**
- AutoML has **clear benefits**:
  - The human is not needed at all, which reduces costs
  - The human can focus on the less menial tasks, which reduces costs and may increase satisfaction
- Why do we need HI then?

HI: Motivation

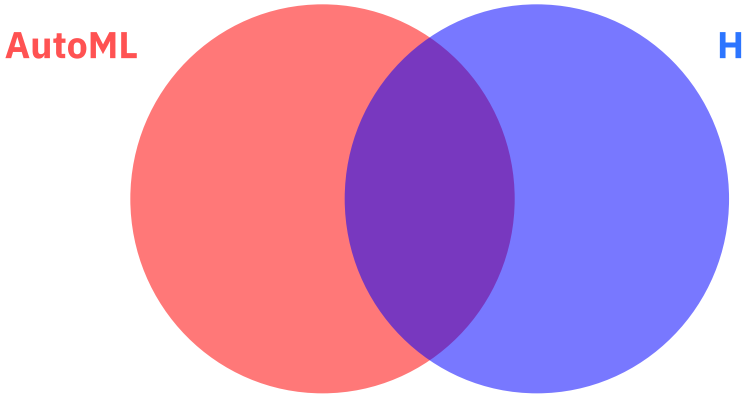
- Some tasks are indeed best suitable for **AutoML**
- Others cannot be solved by AutoML and need **HI**
- Some would see improved results by **combining both approaches**



Ω

HI: Motivation

- Search engine, classification/regression of ontological data
- Visual/multimedia analytics, intelligence on abstract/contextual semantics
- Shared autonomy, long-tail classification/search



Visual/multimedia analytics (HI)

- Task: **expert needs to gain insight into complex data** she hasn't seen before
  - **Forensics**: data relevant to a case, violent online political extremism (VOPE) propaganda...
  - **Medical sciences**: disease-related phenomena on a population (e. g., COVID-19)
  - **Sports**: analytics of plays/tactics employed by various athletes/teams
  - ...
- HI solution: an analytics system with **tightly coupled visualizations and model** that provides on-demand intelligent assistance as the domain expert uncovers "nuggets of insight"

Ω

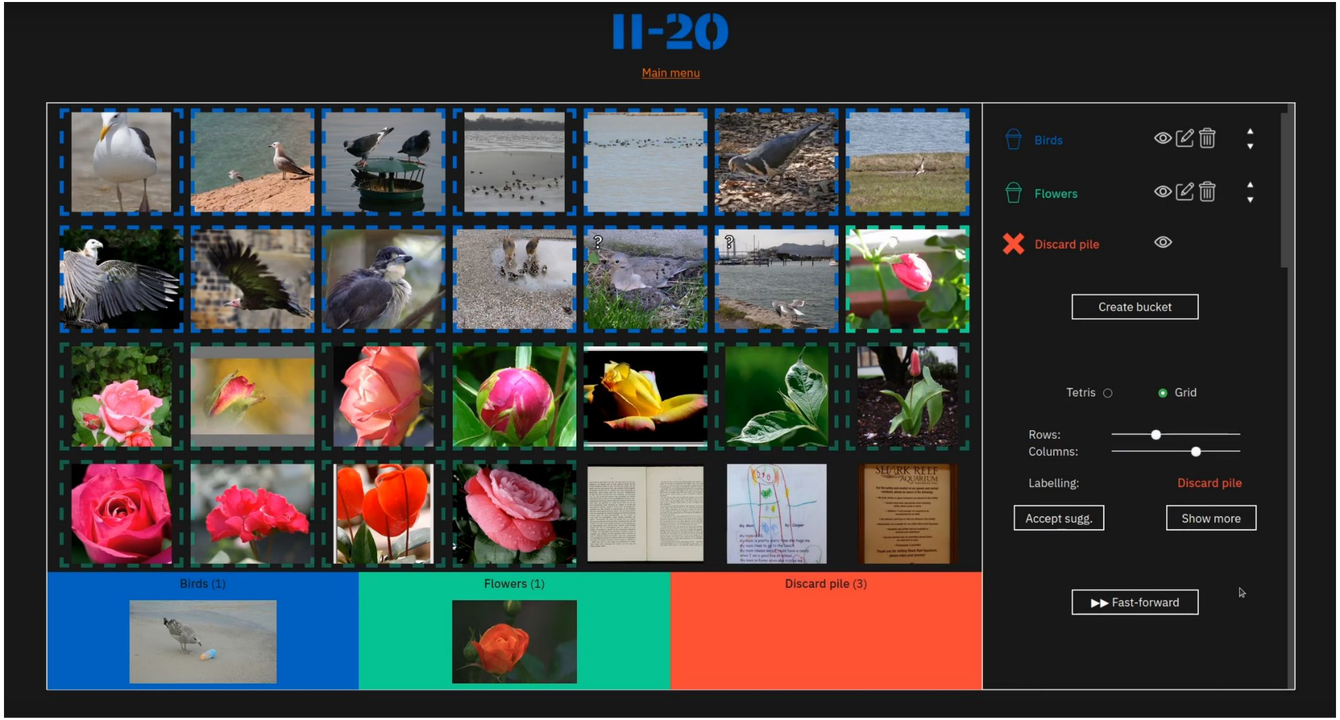
Analytics (HI): Insight

- Insight has the following characteristics [North06]:
  - **Complex** – Insight is complex, involving all or large amounts of the given data in a synergistic way, not simply individual data values.
  - **Deep** – Insight builds up over time, accumulating and building on itself to create depth. Insight often generates further questions and, hence, further insight.
  - **Qualitative** – Insight is not exact, can be uncertain and subjective, and can have multiple levels of resolution.
  - **Unexpected** – Insight is often unpredictable, serendipitous, and creative.
  - **Relevant** – Insight is deeply embedded in the data domain, connecting the data to existing domain knowledge and giving it relevant meaning. It goes beyond dry data analysis, to relevant domain impact.
- Evidently, we need something flexible, interactive

Multimedia analytics (HI): Example

New Yorker Melange - ACM Multimedia Grand Challenge 2014 1st Prize [Zahálka14]

Multimedia analytics (HI): Example



II-20: Intelligent and pragmatic analytic categorization of image collections [Zahálka21]

Abstract/contextual semantics (HI)

- Task: **AI on semantics that are not objective/ontological/predefined**
  - Ad-hoc compounds (`boat on a river`)
  - Feelings/opinions (`beautiful`, `stylish`, `artistic`)
  - Contextual (`suspicious`)
  - Class discovery (`COVID-19 vulnerable`, `XY's new sports tactics`...)
- Injecting user's prior/personal knowledge and intent during the learning process is very important here
- This can be done in HI through a **context dialogue** between the human and the machine

Human vs. machine perception

What we see

$$\begin{bmatrix} (r_{11}, g_{11}, b_{11}) & (r_{12}, g_{12}, b_{12}) & \cdots & (r_{1n}, g_{1n}, b_{1n}) \\ (r_{21}, g_{21}, b_{21}) & (r_{22}, g_{22}, b_{22}) & \cdots & (r_{2n}, g_{2n}, b_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ (r_{m1}, g_{m1}, b_{m1}) & (r_{m2}, g_{m2}, b_{m2}) & \cdots & (r_{mn}, g_{mn}, b_{mn}) \end{bmatrix}$$

What the machine sees

- An m x n (height x width) matrix of pixel RGB values
- This image: m = 3175, n = 4672, so 15.1M values in this one image alone

Image credit: Daniel Chodusov @ Flickr

Semantic gap

- The disproportion between:
  - The information **extractable by a human** from a multimedia item
  - The information **extractable by a machine** from the machine (feature) representation of the same item

$$\begin{bmatrix} (r_{11}, g_{11}, b_{11}) & (r_{12}, g_{12}, b_{12}) & \cdots & (r_{1n}, g_{1n}, b_{1n}) \\ (r_{21}, g_{21}, b_{21}) & (r_{22}, g_{22}, b_{22}) & \cdots & (r_{2n}, g_{2n}, b_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ (r_{m1}, g_{m1}, b_{m1}) & (r_{m2}, g_{m2}, b_{m2}) & \cdots & (r_{mn}, g_{mn}, b_{mn}) \end{bmatrix}$$

Human $\longleftrightarrow$ **Machine**

Complex/abstract semantics
Instant recognition
Put in context

Limited semantics
Takes time, computationally costly
No context

[Smeulders00]

Similarity

- Which image is the odd one out here?

Similarity

- Queen Elizabeth II – the other two both contain a prominent **red** object...

Similarity

- The phone booth – the other two both contain **people**...

Similarity

- Little Red Riding Hood – the other two are both related to **England**...

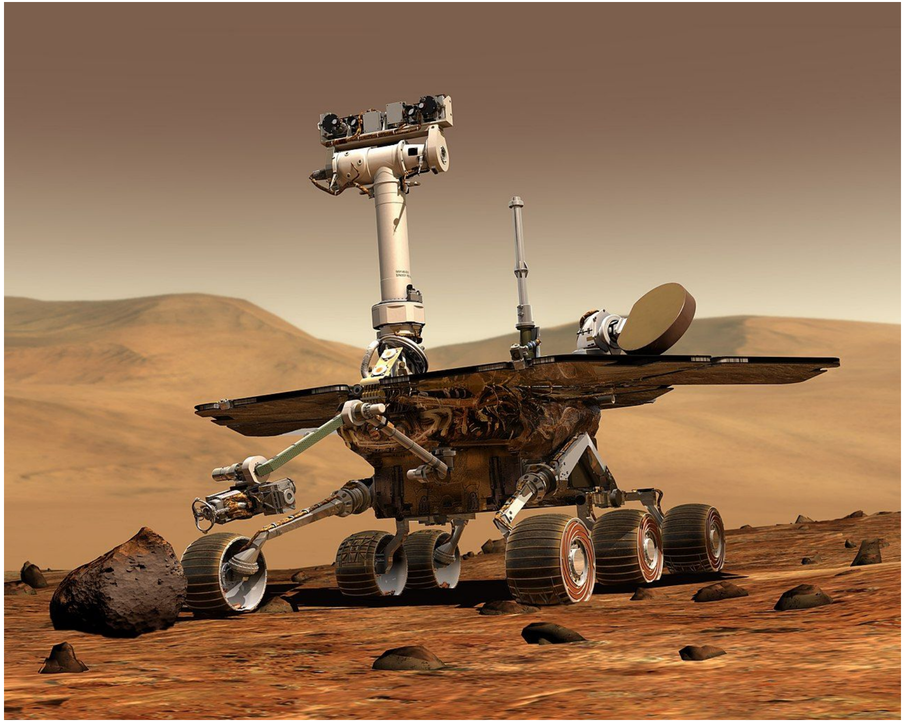
Similarity

- Already on this small collection, one could argue that **structure is not inherent** in the data, but it depends on **context**

Shared autonomy (HI + AutoML)

- Task: **autonomous agent/robot solving a task in an environment**
  - This is generally based on a reinforcement learning model (AutoML)
- **Shared autonomy -** the agent works autonomously, but a human operator can at times intervene and point the agent in the right direction
- Adding HI may result in **faster convergence of the model** and **more trust**, opens up possibility of **"semantic teaching"** of agents/robots
- Could be extended to analytics too
  - For example "write me a report about the data"
  - F. van Harmelen: HI project about creating an HI scientific paper co-author

Shared autonomy (HI + AutoML)

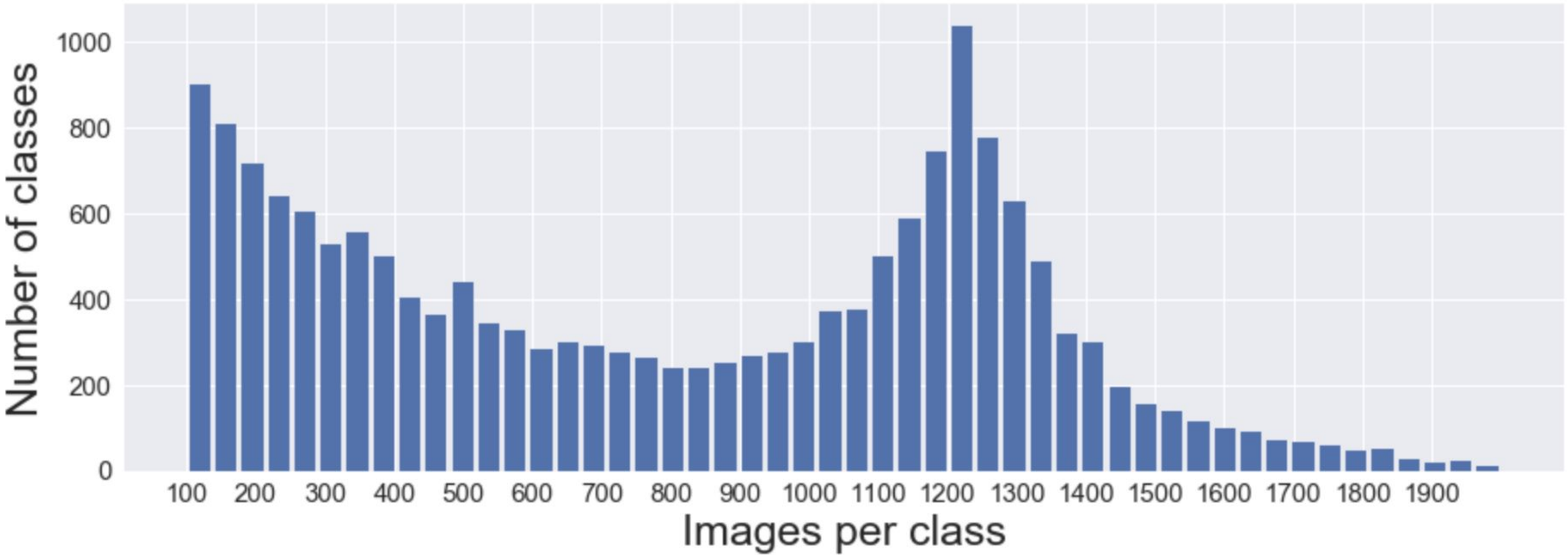


Mars Rover, a pioneer shared autonomy/mixed initiative robot

Long-tailed classification/search (HI + AutoML)

- Task: **meaningful search on semantic concepts that have few/no training examples**
- **Example**: an image database of plants, we're looking for a plant that has only 3 images in the database
- Known phenomenon, great progress in recent years
  - E. g., **zero-shot** & **transfer learning**
- Adding interaction could inject **extra context knowledge** and teach the model to **discriminate the concepts better**

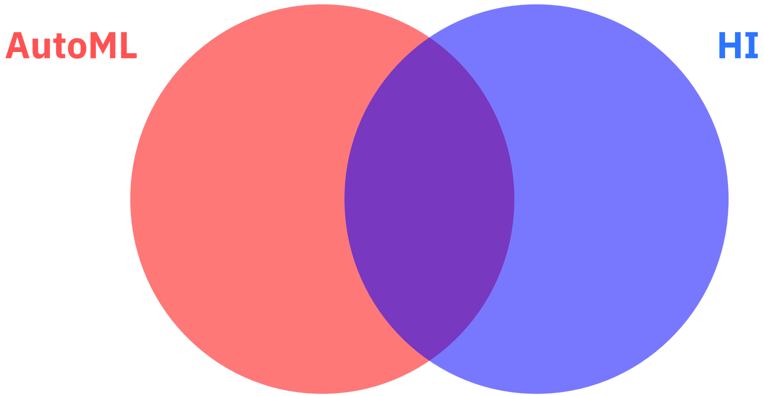
Long-tailed classification/search (HI + AutoML)



ImageNet dataset statistics: thousands of classes with insufficient training data for a conventional deep net

HI: Motivation

- HI **complements** and/or **enhances** AutoML
- Unlocks solving tasks that AutoML wouldn’t solve alone or at all



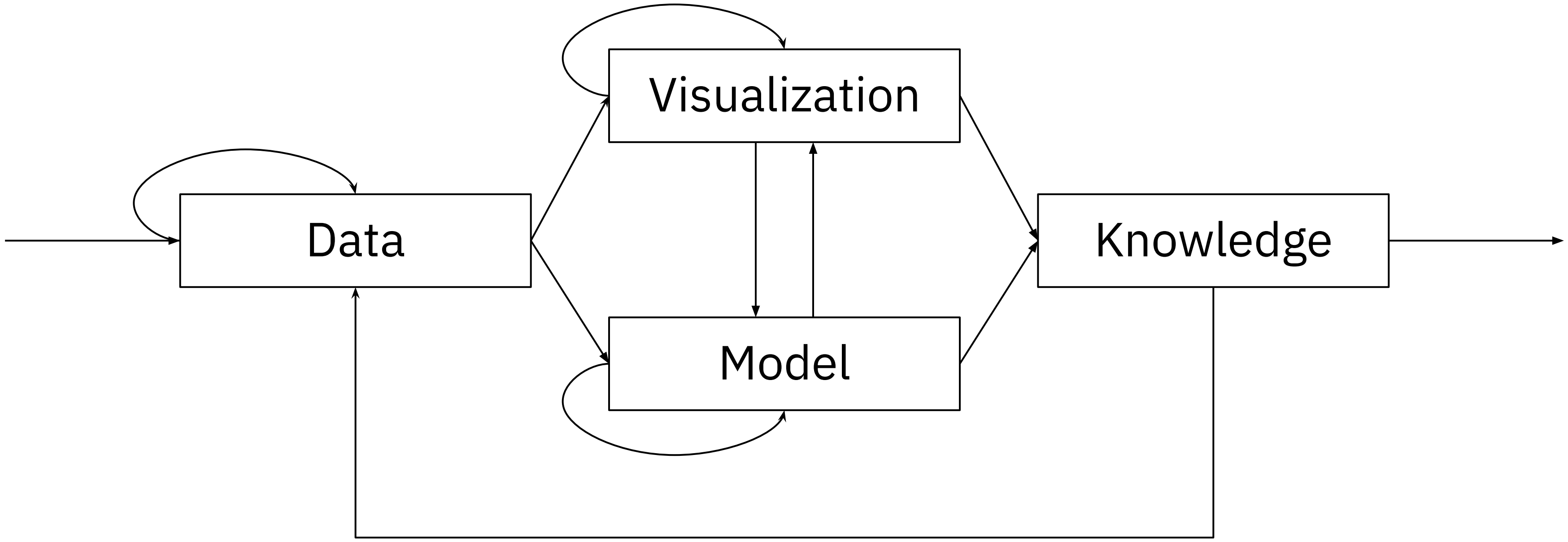
HI benefits: Explainable AI (xAI)

- xAI is a big topic currently, most modern models are black boxes
- Humans want to understand **why a decision was made**
- HI has a degree of **explainability built in**: to successfully lead a dialogue with a human, the interactions that help build the HI model have to make sense

HI benefits: Humanization of AI

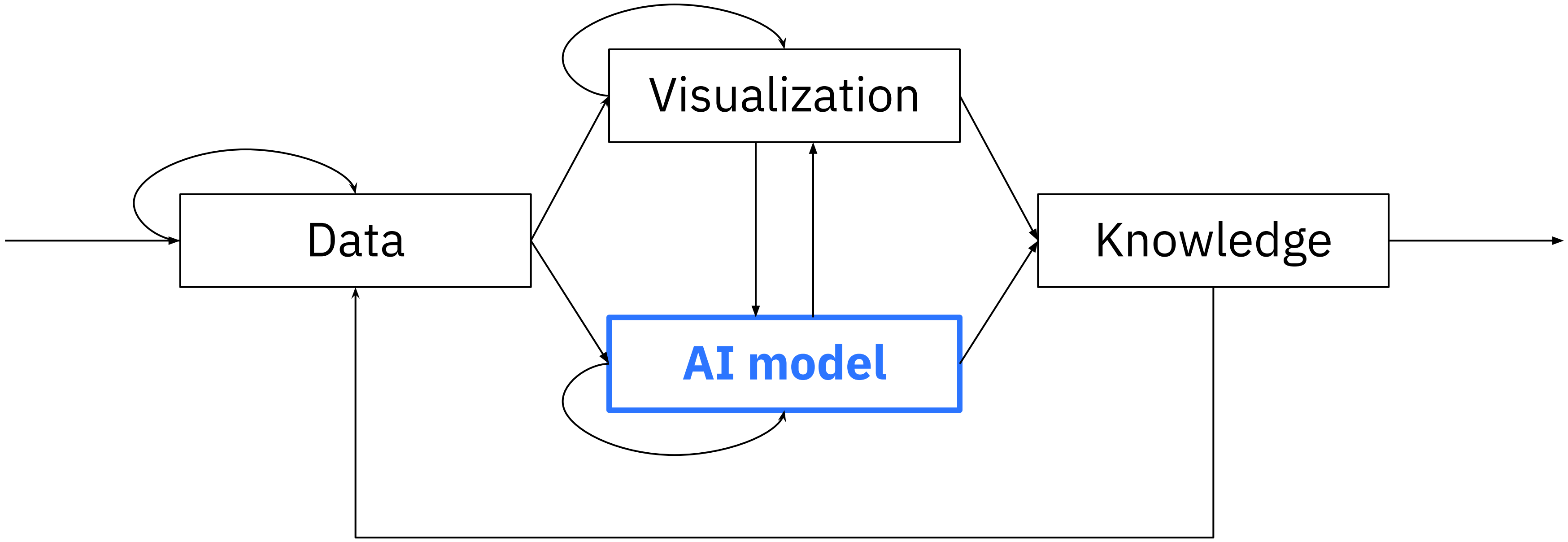
- This is a bit opinionated, but I believe **AI should not dehumanize**
- Considering or treating users as “cattle” that is simply milked for data and then modelled anonymously is **dangerous**
- For specialized tasks with a clear, singular purpose this approach is fine, but as AI becomes more general and generally adopted, it becomes an issue
- HI is centered around the **dialogue between the user and the machine** with the **user’s needs taking priority**

HI conceptualization



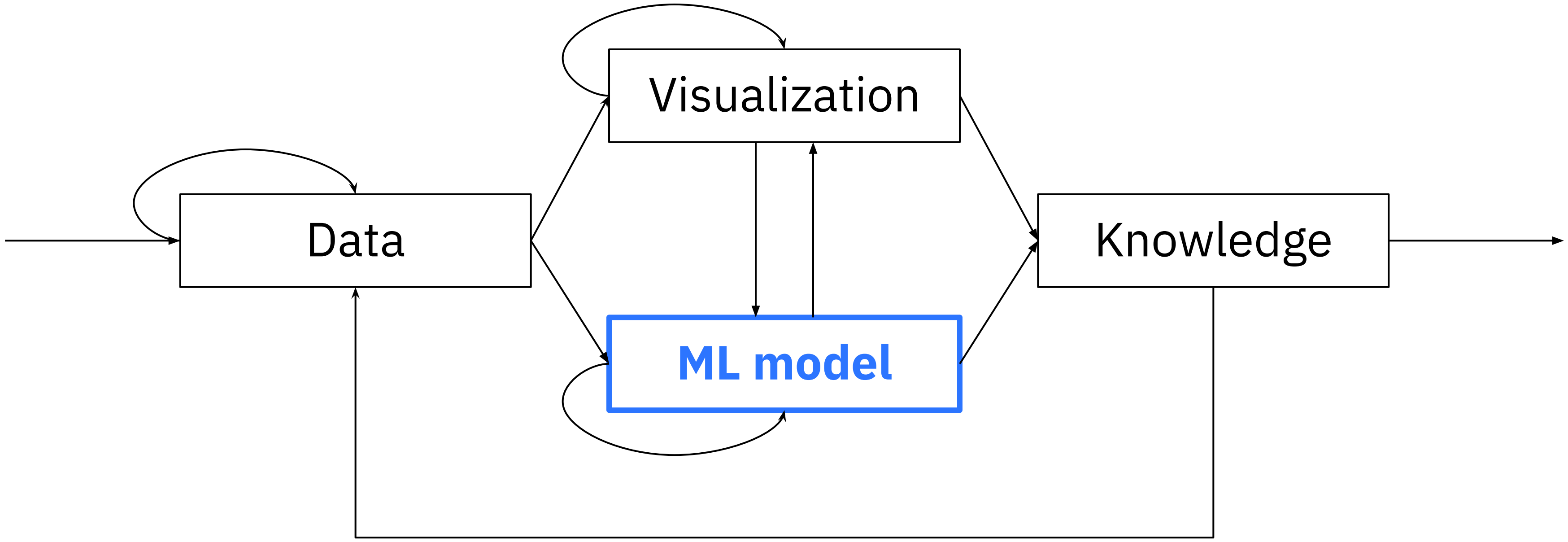
Starting point: Visual analytics pipeline [Keim08]

HI conceptualization



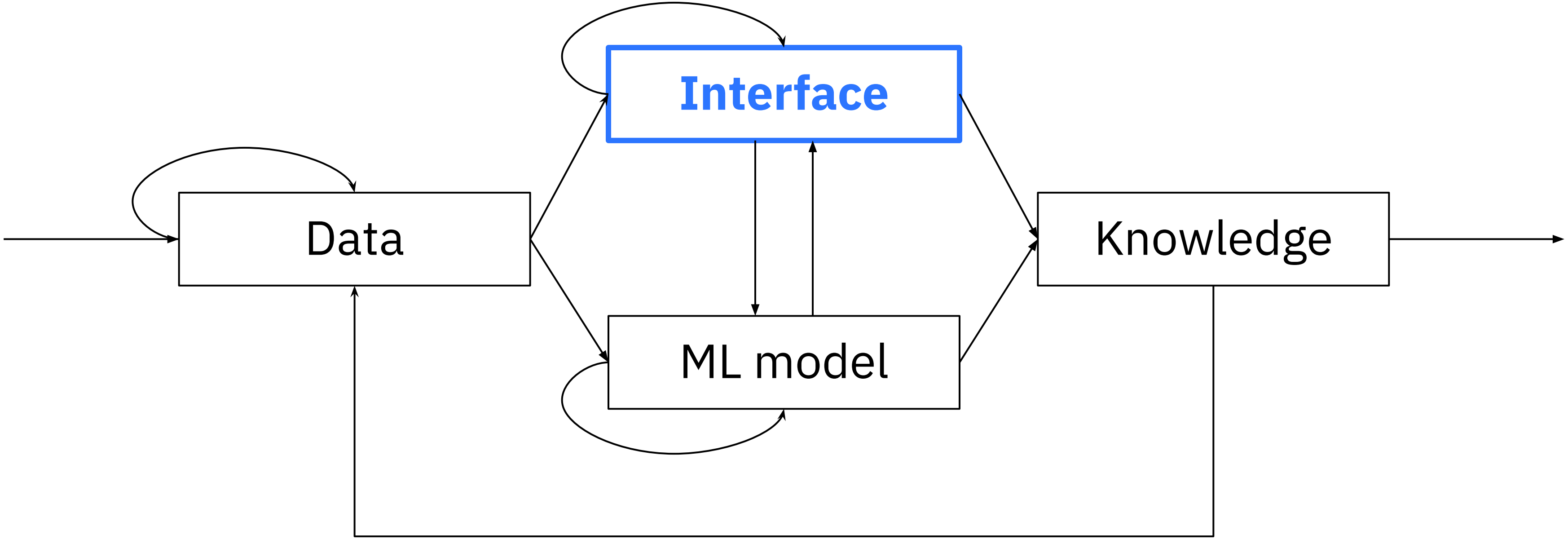
HI = AI + human → the model must be an AI model

HI conceptualization



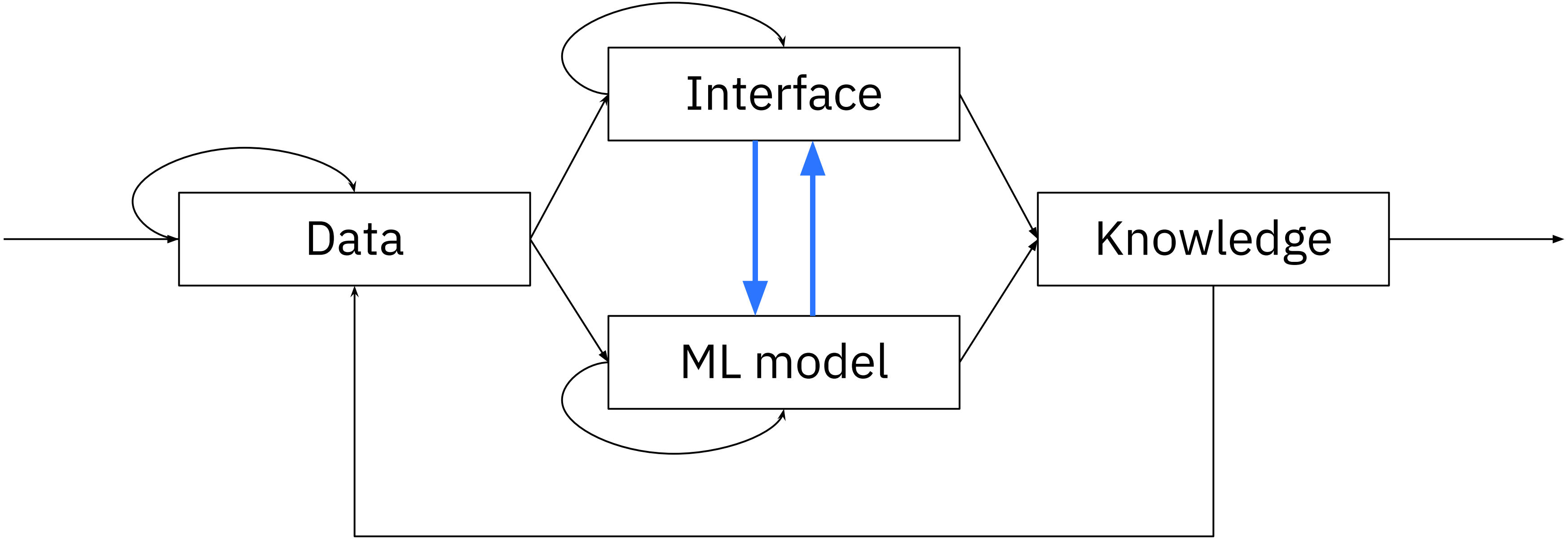
“My flavor” of HI: a (mostly) machine learning model (but again, AI/HI ≠ ML)

HI conceptualization



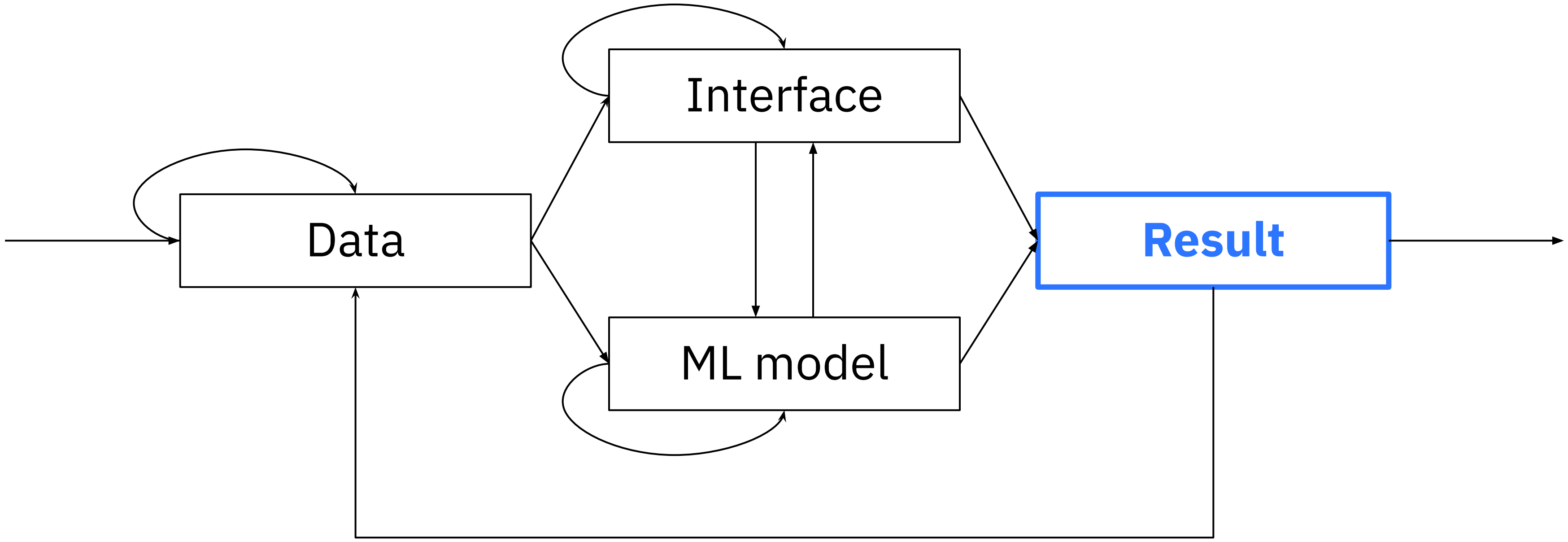
HI does not focus on visualization, but rather has to have a general communication interface

HI conceptualization



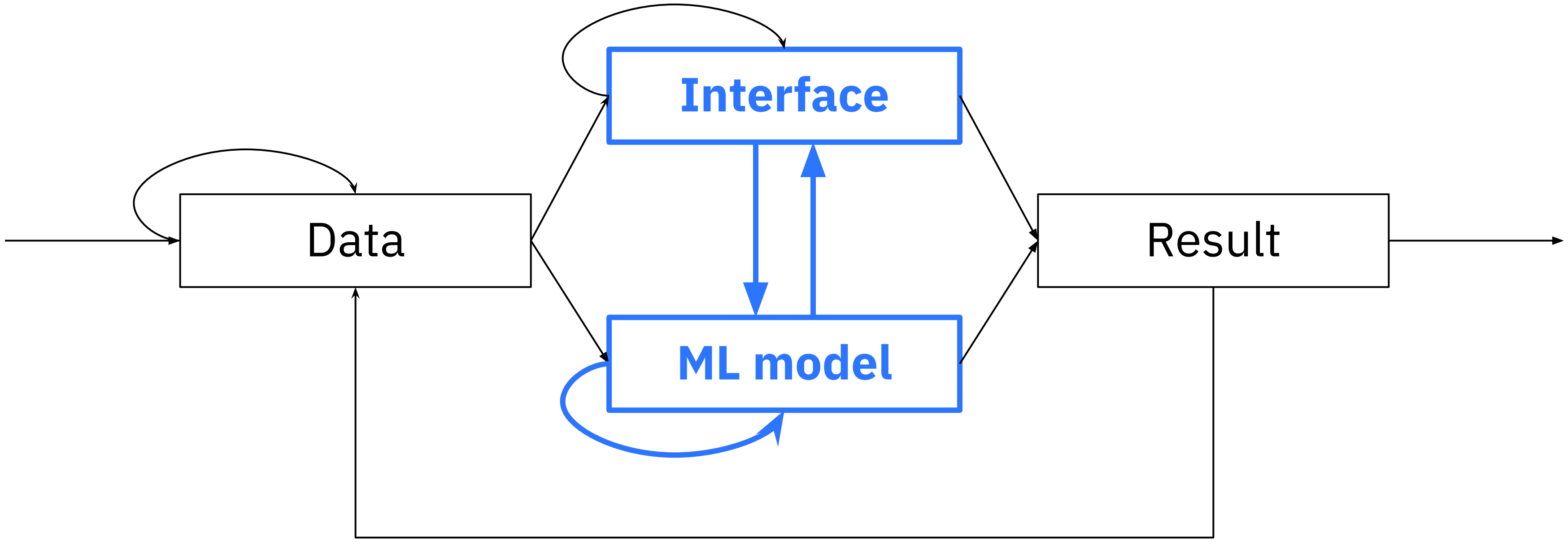
The model must learn from what's communicated and drive further communication

HI conceptualization



HI is not strictly tied to (analytic) knowledge, it can have other results (e. g., complete a certain task)

HI: My focus



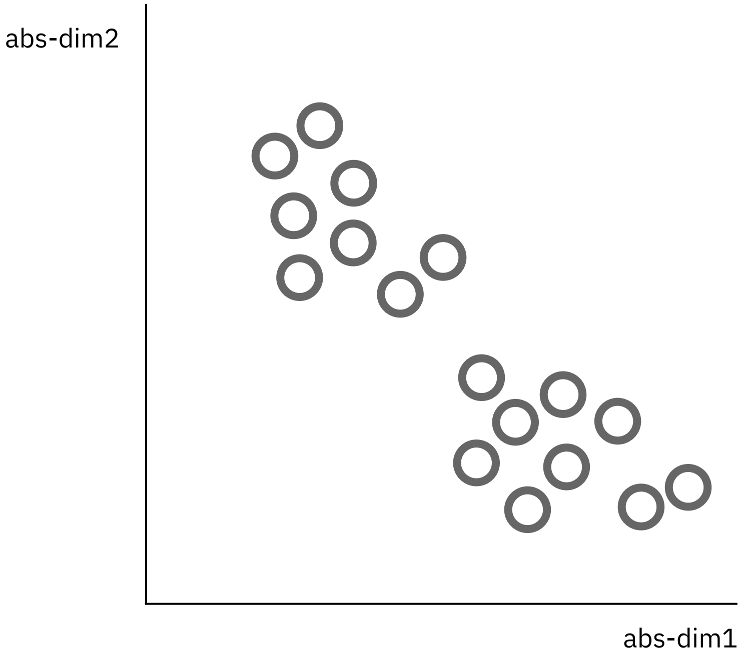
My research focus

Interactive learning (IL)

- A good starting ML vehicle for HI
- Hot in 2000s, fallen out of favour in 2010s mostly due to **deep nets** and **collection size explosion**
- Innovated in late 2010s for **current collection sizes**
- Next focus: **better performance** & **interfacing with AutoML**

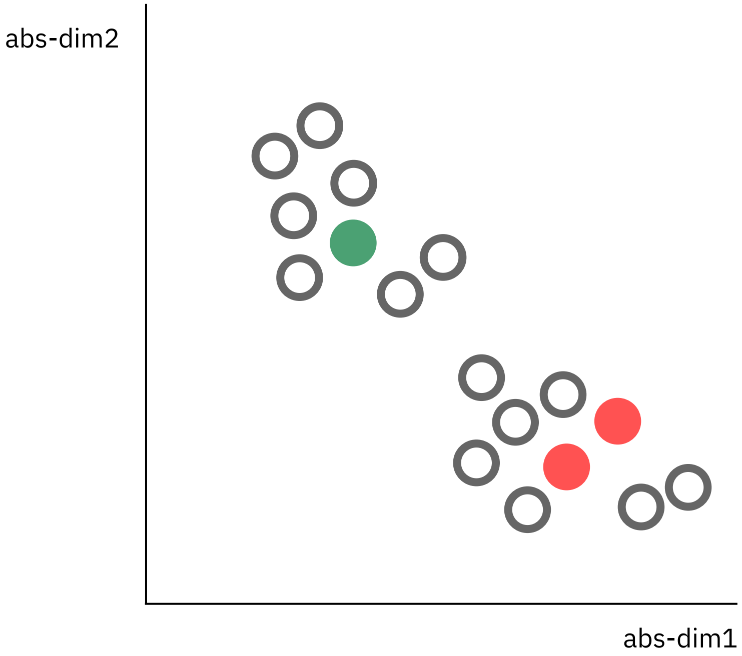
Interactive learning (IL)

1. Extract features first if needed (typically using a deep net)



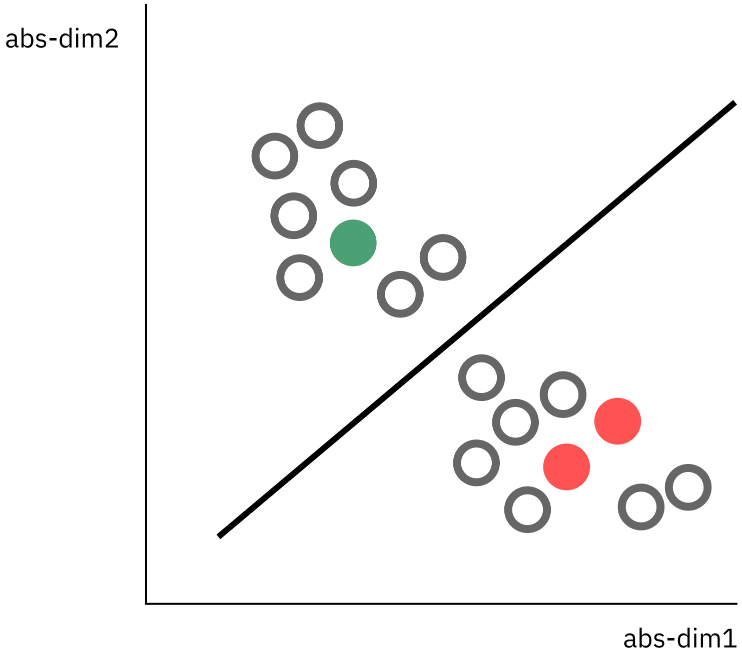
Interactive learning (IL)

1. Extract features first if needed (typically using a deep net)
2. Have a user label a couple of examples as relevant/not relevant



Interactive learning (IL)

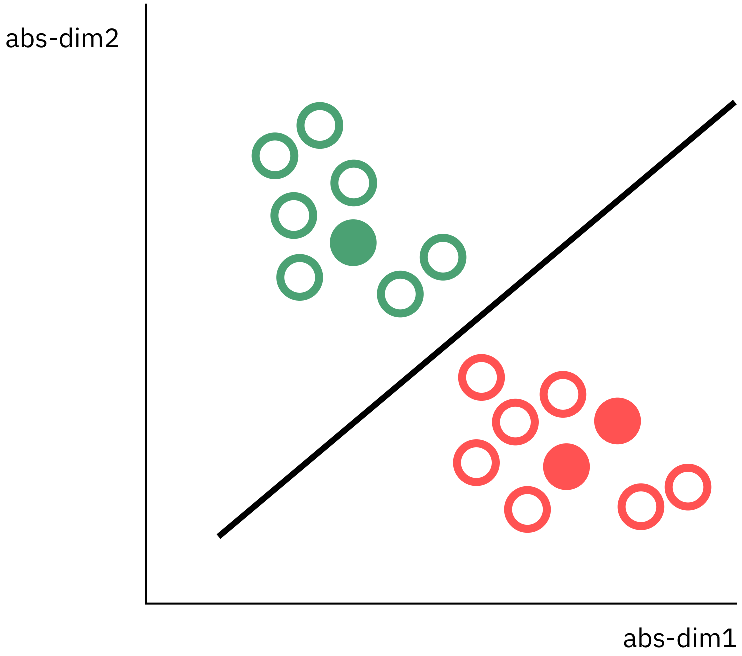
1. Extract features first if needed (typically using a deep net)
2. Have a user label a couple of examples as relevant/not relevant
3. Train the IL model



Ω

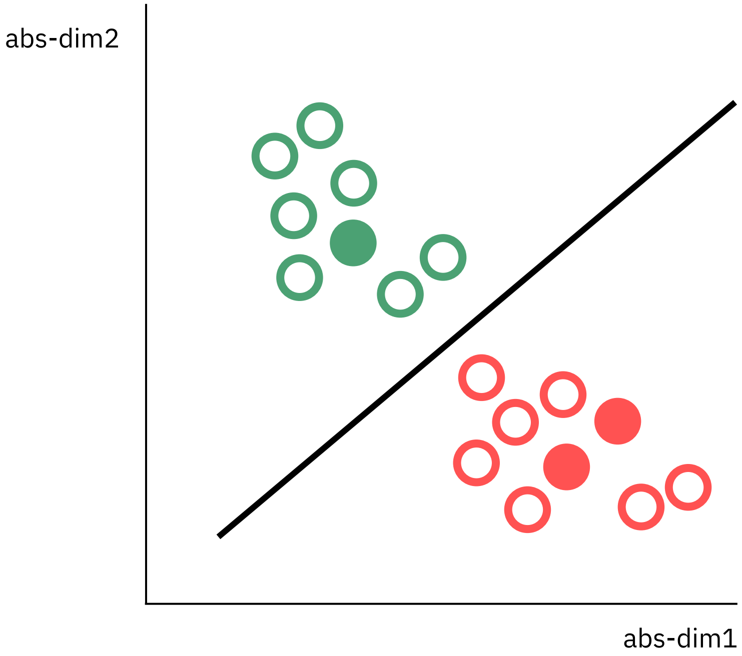
Interactive learning (IL)

1. Extract features first if needed (typically using a deep net)
2. Have a user label a couple of examples as relevant/not relevant
3. Train the IL model
4. Assign class labels & scores to unlabelled data



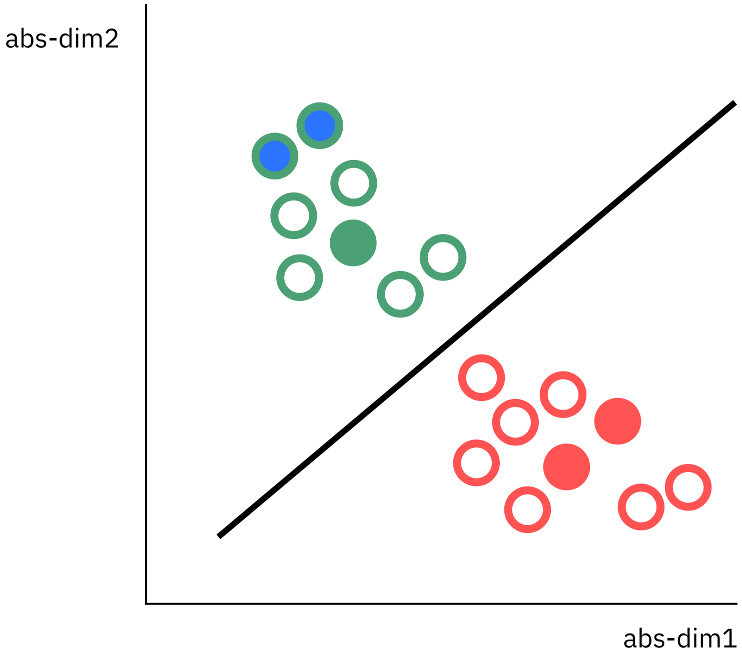
Interactive learning (IL)

1. Extract features first if needed (typically using a deep net)
2. Have a user label a couple of examples as relevant/not relevant
3. Train the IL model
4. Assign class labels & scores to unlabelled data
5. Produce a set of candidates to be labelled by the user in the next round



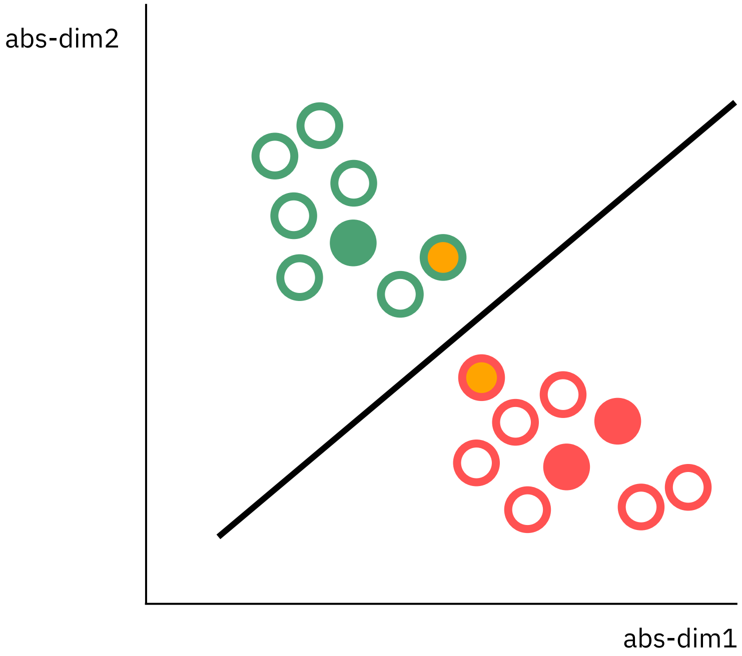
Interactive learning (IL)

1. Extract features first if needed (typically using a deep net)
2. Have a user label a couple of examples as relevant/not relevant
3. Train the IL model
4. Assign class labels & scores to unlabelled data
5. Produce a set of candidates to be labelled by the user in the next round
    - **Relevance feedback** - show the ones where the model is **most confident** (such that the user gets the most likely relevant results)



Interactive learning (IL)

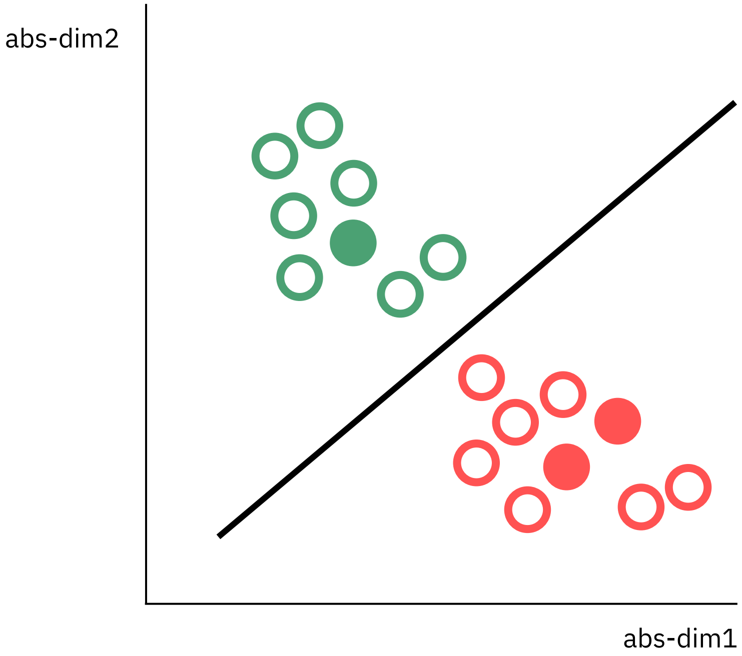
1. Extract features first if needed (typically using a deep net)
2. Have a user label a couple of examples as relevant/not relevant
3. Train the IL model
4. Assign class labels & scores to unlabelled data
5. Produce a set of candidates to be labelled by the user in the next round
   - **Active learning** - show the ones where the model is **least confident** (such that the model converges faster)



Interactive learning (IL)

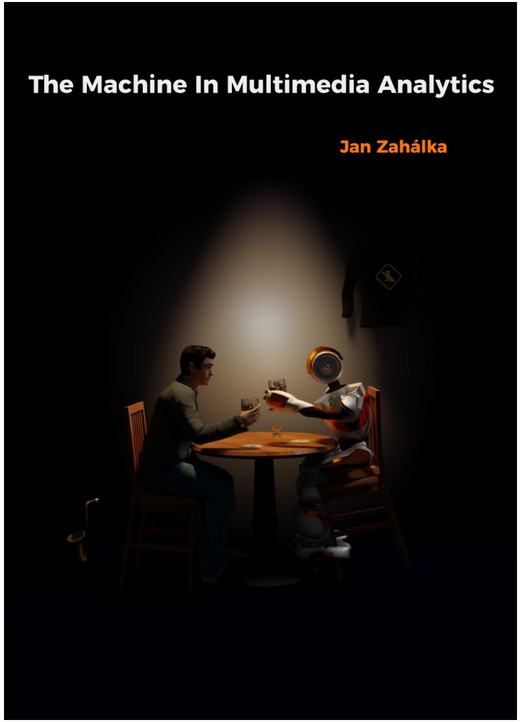
1. Extract features first if needed (typically using a deep net)
2. Have a user label a couple of examples as relevant/not relevant
3. Train the IL model
4. Assign class labels & scores to unlabelled data
5. Produce a set of candidates to be labelled by the user in the next round

Repeat 2-5 until done, steps 3-5 need to complete in **max. 1-2 seconds**



HI: Existing scientific results

- An integrated multimedia analytics **theoretical model**
- Application in **venue recommendation** (New Yorker Melange)
- An **evaluation** framework (Analytic Quality - AQ)
- **Scaling** IL up to 100M images with response time of ~1 s (Blackthorn, Exquisitor)
- A general multimedia analytics **system** for **exploration-search** (II-20)



HI: My research agenda

- **Generalized HI interface** and **richer interaction dictionary** that supports a broad array of HI tasks
  - E. g., such that in analytics it's not only "mark relevant/not relevant items"
  - Currently, you have to write a custom UI/API for each analytics/autonomous agent approach separately, which costs a lot of time

HI: My research agenda

- **New interactive ML algorithms and/or model(s)**
  - The state of the art is pretty much linear SVM on top of features extracted by a deep net
  - Can we do better?
  - How to split offline/online computations?
  - Which model types allow meaningful interplay between each other?

HI: My research agenda

- **Evaluation** - how to evaluate an HI approach esp. in the design phase?
  - User studies are suitable for systems, rather than algorithms/models, and only towards the end of design (or major design cycle)
  - Benchmarks do not take users into account
  - Some work done with AQ [Zahálka15], but more needs to be done

HI: My research agenda

- **Semi-autonomous HI** - the model proceeds with a task autonomously, the human interacts according to the task's needs/their wishes, the model learns from the human and provides a timely response
  - Inspired by shared autonomy, but taken to other tasks such as summarization or active learning

HI: Autonomy axis

**Anything here
is fair game for HI**

Full autonomy

Agents are pretrained, make decisions based solely on the training

Full interactivity

Agents train online on user interactions, solely react intelligently to user

Conclusion

- **Hybrid intelligence** - AI that learns from the user's interactions during the learning/task execution phase
- HI **unlocks AI assistance** in several useful tasks and potentially **enhances** AutoML approaches in others
- Many research opportunities
- Interested? Let's have a chat!
    - `jan.zahalka@cvut.cz`

References

- **[Keim08]** D. Keim et al.: Visual Analytics: Definition, Process, and Challenges. In Information Visualization, Lecture Notes in Computer Science, vol. 4950, Springer, Berlin, 2008.
- **[North06]** C. North: Toward measuring visualization insight. IEEE CGA, 26 (3), pp. 6 – 9, May 2006.
- **[Smeulders00]** A. W. M. Smeulders et al.: Content-based image retrieval at the end of the early years. IEEE PAMI, 22 (12), 1349 – 1380, December 2000.
- **[Zahálka14]** J. Zahálka et al.: New Yorker Melange: Interactive Brew of Personalized Venue Recommendations. ACM Multimedia (ACM MM), pp 205–208, Orlando, USA, November 2014.
- **[Zahálka15]** J. Zahálka et al.: Analytic Quality: Evaluation of Performance and Insight in Multimedia Collection Analysis. ACM Multimedia (ACM MM), pp. 231–240, Brisbane, Australia, October 2015.
- **[Zahálka21]** J. Zahálka et al.: II-20: Intelligent and pragmatic analytic categorization of image collections. IEEE TVCG, 27 (2), pp. 422 – 431, February 2021.