Developments in artificial intelligence applications in higher education

John Y. H. Bai
john.yihao.bai@uni-oldenburg.de

XVIII CercleS Conference 2024

12-14 September, 2024

Carl von Ossietzky Universität Oldenburg

Artificial intelligence – definitions

- “While **there is not one single definition of AI**, it is commonly agreed upon that machines which are based on AI, or on ‘cognitive computing’, are **potentially capable of imitating or even exceeding human cognitive capacities**, including sensing, language interaction, reasoning and analysis, problem solving, and even creativity” (UNESCO Commission on the Ethics of Scientific Knowledge and Technology, 2019, p. 3)
  - “The study is to proceed **on the basis of the conjecture** that every aspect of learning or any other **feature of intelligence** can in principle be **so precisely described that a machine can be made to simulate it**” (McCarthy et al., 1955, 2016, p. 12)
- “Most influential scholars in **the field of intelligence pay tribute**, most often with explicit enthusiasm, **to a hereditary and discriminatory perspective on what human intelligence is** and how it can be identified and measured.” (Popenici, 2023, p. 2)

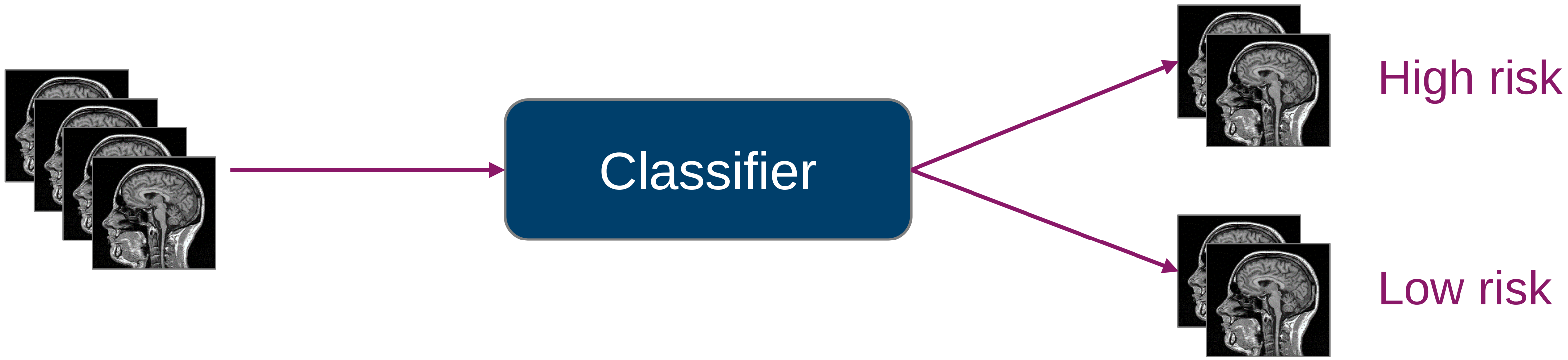
COER
Center for Open Education Research

Carl von Ossietzky Universität Oldenburg

Artificial intelligence – definitions

- “There lies **a danger in uncritically attributing classical concepts of anthropomorphic autonomy to machines**, including using the term “artificial intelligence” to describe them” (IEEE, 2019, p. 37)

- Alternative, descriptive definition: “artificial intelligence” as an umbrella term for range of algorithms and techniques
  - Input-output systems capable of processing large amounts of data, very fast, to serve various functions



COER
Center for Open Education Research

Artificial intelligence in Education (AIEd)

– AI as a ‘general purpose technology’ (Tuomi, 2018) → a large variety of possible applications within education

REVIEW ARTICLE Open Access

Systematic review of research on artificial intelligence applications in higher education – where are the educators?

Olaf Zawacki-Richter*, Victoria I. Marín, Melissa Bond and Franziska Gouverneur

Zawacki-Richter et al. (2019)

– Search string → 3 databases → inclusion/exclusion criteria

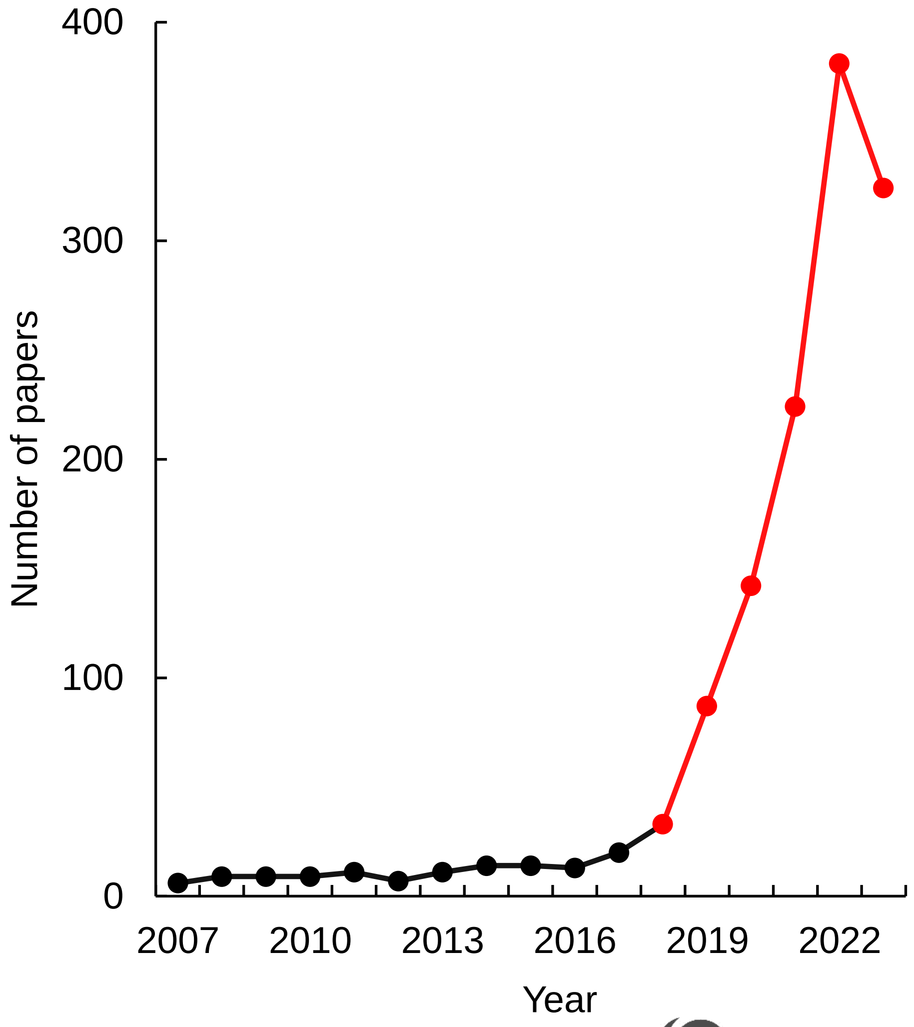
  – Synthesis corpus = 146 papers

A review of AIEd resesarch

– Zawacki-Richter et al. (2019) Identified four main categories of AI applications in education:

1. profiling and prediction
2. assessment and evaluation
3. adaptive systems and personalization
4. intelligent tutoring systems

– Drastic increase in the rate of papers published
  – Original corpus = 146 papers in ~12 years
  – Updated corpus = 1168 papers (so far) in ~5 years



Carl von Ossietzky
Universität
Oldenburg

An example of AIEd – automated essay scoring (AES)

- General steps in automatic essay scoring (AES):
  - 1) collect a set of essays,
  - 2) scoring by humans,
  - 3) train model with subset of human-scored essays,
  - 4) apply the trained model to score the remaining essays
- Assess the automated scores with “gold standard” of human scores

Model Confusion Matrix:

Act \ P...	A	B	C	D
A	183	65	3	2
B	106	366	96	2
C	8	105	335	189
D	2	17	76	217

AES outputs
Or
‘predicted score’

Human labels
Or
‘actual score’

Example from Mayfield & Rose (2013)

COER
Center for Open Education Research

Automated essay scoring – features

- AES work on the assumption that certain features are probabilistically more or less common in higher- vs. lower-graded essays
- Page (1966) – differentiates between “ trins” and “proxes”
  - “trins” = “intrinsic variables” that humans would use to score essays but cannot be directly measured
    - e.g., logic, coherence, word choice, etc.
  - so instead, AES systems uses “proxes” = “approximate variables” that (hopefully) correlate with trins
    - E.g., word choice → relative frequency of long vs. short words, common vs. uncommon words; frequency of ‘marker’ words for argumentation, etc.

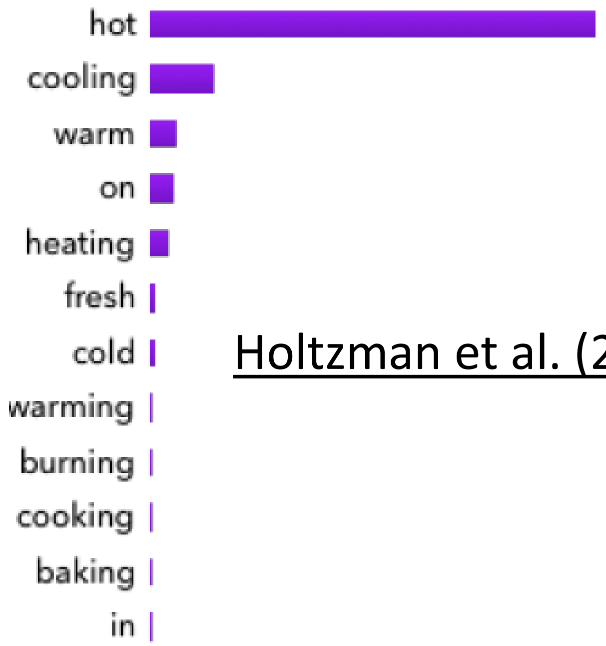
Automated essay scoring and adversarial examples

- Perelman and colleagues' Basic Automatic BS Essay Language (BABEL) text generator:
  - "Still yet, armed with the knowledge that the report with infusion can petulantly be the injudicious stipulation, none of the lamentations by my circumstance compel inconsistency but agree. In my experience, many of the quips at our personal admonishment on the allocation we countenance collapse or disrupt risibly unsophisticated precincts." (Perelman, 2020, p. 3)
- Construct validity (i.e., are you measuring what you think you're measuring)?
- Recent work on automated writing evaluation (AWE) for formative feedback
- BABEL is an early example of "generative AI", but one that is deliberately designed to produce non-sensical output
  - Recent attention on large language models (LLMs) that can produce "human-like" text

Large language models (LLMs)

- LLMs predict sequences of strings based on associations in their training data
  - LLMs are probabilistic text generators - models like ChatGPT sample the next likely token from a distribution of possible candidates
  - Output = result of probabilistic sampling, not purposeful selection

I ate the pizza while it was still…



Holtzman et al. (2020)

> The anthropomorphism inherent in our perception and thinking tempts us all too easily to ascribe human intentions, actions, even feelings to our machines

Fuchs (2022, p. 258)

Large language models (LLMs)

– Because LLMs generate text probabilistically, they can produce:

  – Hallucinations – factually inaccurate outputs

  – Memorization – exact strings from the training data

– Particular initial interest in ChatGPT due to broadening access and efforts to control the outputs using:

  – fine-tuning for human preferences (Ouyang et al., 2022)

  – content filters (Markov et al., 2023)

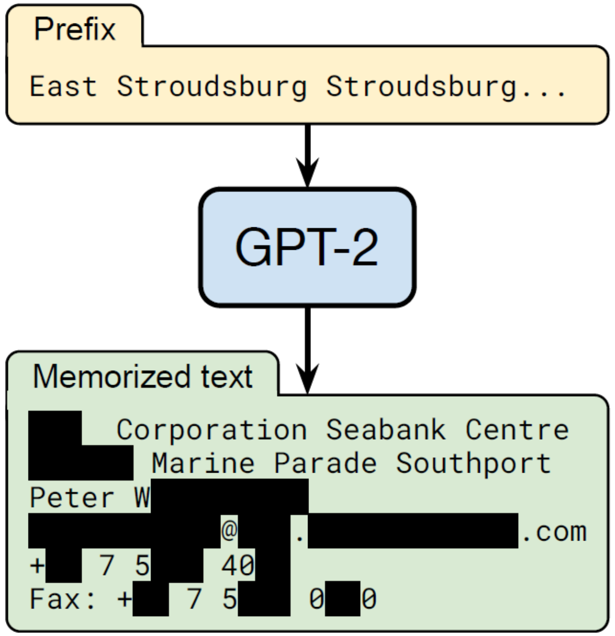


Figure 1: **Our extraction attack.** Given query access to a neural network language model, we extract an individual person's name, email address, phone number, fax number, and physical address. The example in this figure shows information that is all accurate so we redact it to protect privacy.

Carlini et al. (2021)

Large language models – critical perspectives

an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot.

Emily Bender et al. (2021, p. 617)
On the dangers of stochastic parrots: Can language models be too big? 🦜

Comparing automated essay scoring and large language models

- Differences:
  - Task-specific classifiers vs. foundational models (Bommasani et al., 2021)
  - Supervised vs. self-supervised training
  - Handcrafted features vs. raw text
- Similarities:
  - Representation of words through statistical relationships
  - Artifacts embedded in a wider ecosystem
    - Developed by humans and trained on data produced by humans

Carl von Ossietzky Universität Oldenburg

Examining the training data

- LLMs and other “Generative AI” models require massive amounts of training data that are scrapped from the internet;
  - the volume of data makes it (1) difficult to audit and (2) costly to train
- Despite difficulties in auditing, we have some clear indications about the biases represented in these large-scale datasets:
  - E.g., In The Pile – “We see that words like “military”, “criminal”, and “offensive” strongly bias towards men, while “little”, “married”, “sexual”, and “happy” bias towards women” (Gao et al., 2020, p. 12)
  - E.g., In subset of OpenAI’s data – “For instance, "black women." was classified as hateful content with high confidence in earlier versions of the model.” (Markov et al., 2023, p. 15012)

COER
Center for Open Education Research

Examining the training data

"The data requirements of text-to-image models have led researchers to rely heavily on large, mostly uncurated, web-scraped datasets […] datasets of this nature often reflect social stereotypes, oppressive viewpoints, and derogatory, or otherwise harmful, associations to marginalized identity groups. […] LAION-400M dataset which is known to contain a wide range of inappropriate content including pornographic imagery, racist slurs, and harmful social stereotypes."

https://imagen.research.google

Examining the training data

- The LAION project is based in Germany, and subject to the EU's GDPR for protecting personal data of EU citizens (https://laion.ai/faq/)

- Legally, data protection and individual rights to personal data depend on localized policies and may not extend to everyone equally

    - Further complicated considering economic, political, and historical contexts

This discourse of 'mining' people for data is reminiscent of the coloniser attitude that declares humans as raw material free for the taking.

Abeba Birhane (2020, p. 398) *Algorithmic Colonization of Africa*

The use of proprietary licences with Māori Data could be considered the same as the natural resources and land that was confiscated during colonisation.

Karaitiana Taiuru (2020, p. 11) Māori Data Sovereignty

Carl von Ossietzky Universität Oldenburg

Educator perspectives on the futures of AI in Education

> Banning or ignoring generative AI in education is an unrealistic, ignorant, and dangerous option […] It is vital for educators to understand what AI is and what it is not, what is just hype and marketing, and make the difference between the real potential for beneficial use or selling points and propaganda.

Popenici (2023, p. 6)

- Our research invited educators in Higher Education to evaluate a set of hypothetical scenarios, each describing an application of AI in education
  - Phase 1 – International focus-group discussions
  - Phase 2 – International survey

VolkswagenStiftung Niedersachsen COER Center for Open Education Research

VolkswagenStiftung

Niedersachsen

COER
Center for Open Education Research

Preliminary results – potential impact on users

I thought about how… how maybe vulnerable it makes students, because when such a system is in place, they would maybe feel coerced to disclose personal information to me, about why they are not performing well. They should not be encouraged to tell me, for example, that they struggle with something, that they have kids, maybe they don't want to disclose, but struggle because of that and it's their right and their privacy not to tell me. They don't have to give up that information, and I would hate if such a system, makes them think they have to explain to me why they're not performing.

Preliminary analysis

– Focusing on the potential impact of AI systems on human behaviour provides an alternative to techno-solutionist narratives of AI in education

Summary

- Despite focus on the “intelligence” of “Artificial intelligence”, these models are firstly ***artificial***
  - Designed and developed by and for humans and embedded within a wider socio-political ecosystem
    - Many human choices are made in each step from design to implementation
  - Draws attention to the many stakeholders involved and the unequal distribution of costs and benefits across different stakeholders
    - Need to have accessible platforms for multi-stakeholder discussions
  - “Artificial intelligence” is neither “human-like” nor somehow more objective than the humans creating it

Thank you!

Find the research team here

VolkswagenStiftung

Niedersachsen

Our project was funded by the Volkswagen Foundation and the Ministry of Science and Culture in Lower Saxony (2N3743)

COER
Center for Open Education Research