Communicative AI in Education – from bias to trust

Prof. Dr. Andreas Breiter

AI and Data use in Education (AID-E) network
University of Twente

16.5.2023
Today’s agenda

• AI in Education – mapping the field
• Communicative AI and socio-technical AI risks
• Bias ... in AI
• Potential solutions
• Conclusions
Can AI Do Teacher Observations and Deliver PD? In Some Schools, It Already Does

New York City blocks use of the ChatGPT bot in its schools

ChatGPT and the future of university assessment

Artificial intelligence-powered tools like ChatGPT are forcing a much-needed opportunity to reimagine the role of education in the 21st century, says Alex Sims
AI = Advanced high-tech for reoccurring goals in education

„improving students’ learning experience, promote personalisation and better learning outcome“

... for sustainable development

AI in Education – Mapping the field*

Learning and Teaching
• Intelligent Tutoring Systems
• Assessment,
• Feedback,...
• Personalisation!

Learning Management
• Course recommendation
• Learning analytics
• Retention prediction, ...
• Optimization!

Student services
• Admission,
• Advising,
• Well-being,...
• Rationalization!

Generic AI: Chatbots, NLP, Large-language models, image/audio/video processing

*It is not new!
• 1989: International Journal of Artificial Intelligence in Education
• 1993: International AI in Education Society (IAIED)
Publications have increased exponentially

Documents by year

Source: Scopus.com
Search term: “AI AND in AND education”
Research studies have increased, too

Analysis of data of 138 studies from Higher Education (2016-2022)

In which subject domains in HE are AI systems studied?

(2016-2022)

# Examples of AI use at the University of California

<table>
<thead>
<tr>
<th>Admissions and Financial Aid</th>
<th>Retention, Student Advising, and Academic Progress</th>
<th>Student Mental Health and Wellness</th>
<th>Grading and Remote Proctoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>• AI programs are available to improve admissions yield ... and reduce the summer pre-enrollment gap or “summer melt”</td>
<td>• AI programs can provide feedback on student engagement and performance, assist academic advisors, and ultimately improve student retention, academic success, and graduation rates.</td>
<td>• AI can help attend to the social-emotional needs of students during this critical time of their early adulthood and flag cases for human intervention as warranted.</td>
<td>• AI can be used to automate grading, detect academic misconduct, and proctor tests remotely.</td>
</tr>
</tbody>
</table>

What is new in AI ... and in Education?

• Advances in Machine Learning (especially maturity of Deep Learning Approaches)
  ?? “Perceptrons” invented in 1957 (Rosenblatt, 1957, Minsky & Papert 1969)

• Faster algorithms, better hardware for data processing
  ➔ Moore’s Law

• Availability of Data, as in Large Language Models for Generative AI
  ?? “Stochastic parrots” (Bender et al., 2021)
Today’s focus: Communicative AI

• Automation of communication – communicative automation
  • How does communication change when AI becomes part of it?
  • “Communicative AI technologies are not only designed to function as communicators but are also interpreted by people as such.” (Guzman, 2019)
• Part of a societal meta-trend of “datafication” in education (Jarke & Breiter, 2019)
IMPACT Project

Implementing AI-based Feedback and Assessment with Trusted Learning Analytics in Higher Education

• 5 German Universities
• 200,000 students
• 4 years (2020-2024)
• Implementation study
• ELSI perspectives with SHEILA framework

- ComAI for Q&A (internships, teacher training, General Studies)
- Learning Analytics and ComAI for formative assessment and feedback (Blogs, videos, short essays)
- Learning Analytics for summative assessment and feedback (essays, exams)
CommAI in IMPACT Project

1. General purpose chatbot (intelligent Q&A)

2. Formative feedback chatbots

**Biology:** Intro to Statistics
- Formulas
- Graphs
- Text explanations

**Information Systems:**
- Intro to Service Management
- Explanatory Videos
  - Feature extraction (YOLO)
  - Text analysis

Domain-specific chatbots (RASA)  LLM??
“... data does not just exist — it has to be generated. Data creators have to collect data and organize it, or create it from scratch. Texts need to be written, photographs need to be taken, video and audio material need to be recorded. Or they need to be digitized from already existing media.” (p.224)

“Rather, data are generated: produced through measuring, abstracting and generalising techniques that have been conceived to perform a task and are recorded into forms and measures that conform with standards invented by people.” (p.15)


Bias in Machine Learning

Support The Guardian

Who won and who lost: when A-levels meet the algorithm

© Twitter (@teenybisquit)
Types of bias in machine learning pipeline

Bias in education?
For a historical reflection, see Hutchinson (2019)

Examples for Bias from project experience

• Historical:
  • Student demographics as predictors for grades (e.g. perpetuating historical differences between different gender)

• Representational:
  • Under-sampling a group in the training data sets (e.g. native/non-native speakers; students with disabilities)

• Measurement
  • Variables without construct validity (e.g. inaccuracies due to changes in the tasks/courses/learning units from semester to semester: this can result in incorrect indicators or the lack of feedback texts)
Examples for Bias from project experience

• Aggregation
  • Combining distinct populations in the same model, which does not work for some (or all) groups (e.g. detectors of retention trained on a combination of students from MINT, Social Science and Teacher Education, which functions poorly for all three groups than detectors trained on individual groups).

• Evaluation
  • Using test sets to evaluate a model which do not represent the population on which the model should work (e.g. test set with first year students for all students).

• Deployment
  • Using the models in ways they where not originally designed for (e.g. a model designed to identify student disengagement for formative assessment which is later used to assign grades.)
Potential Solutions
Explainable AI – opening the „black box“

• “Explicability turn” (Farrow 2023)
• Ethical frameworks, extensive research and many suggestions
• 47 proposed principles boiled down to four traditional ethical principles
  • beneficence; non-maleficence; autonomy; and justice
  • explicability (Floridi, Cowls, and Beltrametti 2018; Floridi and Cowls 2019)
• „Solving tech problems with tech“
### Challenges for explainable AI

<table>
<thead>
<tr>
<th>Confidentiality</th>
<th>An algorithm may be confidential for reasons of competitive edge or trade secrecy; or as a matter of public security</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td>Some algorithms are clearly understood by experts, but their complexity cannot easily be communicated to the layperson</td>
</tr>
<tr>
<td>Unreasonableness</td>
<td>Algorithms might use rationally justifiable decisions to implement decisions or actions which are unfair or discriminatory</td>
</tr>
<tr>
<td>Injustice</td>
<td>Algorithms may be understood in their operation but the legal and/or moral consequences also need to be explicated</td>
</tr>
</tbody>
</table>

UC RESPONSIBLE AI PRINCIPLES

In response to growing concerns over the use and consequences of AI, at least 170 sets of AI principles and guidelines have been developed to guide the private and public sectors’ responsible development and use of AI.19 While the sets of principles vary in style and scope, a consensus is growing around key themes, including the need for accountability, privacy and security, transparency and explainability, fairness and non-discrimination, professional responsibility, human control, and the promotion of human values like civil and human rights.20

Drawing upon these common themes and from insights gleaned from the deep expertise of the members of the UC Presidential Working Group on Artificial Intelligence, the Working Group urges UC to adopt the following responsible AI principles to guide its procurement, development, implementation, and monitoring of AI within its provision of services:

1. **Appropriateness**: The potential benefits and risks of AI and the needs and priorities of those affected should be carefully evaluated to determine whether AI should be applied or prohibited.

2. **Transparency**: Individuals should be informed when AI-enabled tools are being used. The methods should be explainable, to the extent possible, and individuals should be able to understand AI-based outcomes, ways to challenge them, and meaningful remedies to address any harms caused.

3. **Accuracy, Reliability, and Safety**: AI-enabled tools should be effective, accurate, and reliable for the intended use and verifiably safe and secure throughout their lifetime.

4. **Fairness and Non-Discrimination**: AI-enabled tools should be assessed for bias and discrimination. Procedures should be put in place to proactively identify, mitigate, and remedy these harms.

5. **Privacy and Security**: AI-enabled tools should be designed in ways that maximize privacy and security of persons and personal data.

6. **Human Values**: AI-enabled tools should be developed and used in ways that support the ideals of human values, such as human agency and dignity, and respect for civil and human rights. Adherence to civil rights laws and human rights principles must be examined in consideration of AI adoption where rights could be violated.

7. **Shared Benefit and Prosperity**: AI-enabled tools should be inclusive and promote equitable benefits (e.g., social, economic, environmental) for all.

8. **Accountability**: The University of California should be held accountable for its development and use of AI systems in service provision in line with the above principles.
RESPONSIBLE ARTIFICIAL INTELLIGENCE

UC RESPONSIBLE AI PRINCIPLES

In response to growing concerns over the use and consequences of AI, at least 170 sets of AI principles and guidelines have been developed to guide the private and public sectors’ responsible development and use of AI. While the sets of principles vary in style and scope, a consensus is growing around key themes, including the need for accountability, privacy and security, transparency and explainability, fairness and non-discrimination, professional responsibility, human control, and the promotion of human values like civil and human rights.

Drawing upon these common themes and from insights gleaned from the deep expertise of the members of the UC Presidential Working Group on Artificial Intelligence, the Working Group urges

4. Fairness and Non-Discrimination: AI-enabled tools should be assessed for bias and discrimination. Procedures should be put in place to proactively identify, mitigate, and remedy these harms.

6. Human Values: AI-enabled tools should be developed and used in ways that support the ideals of human values, such as human agency and dignity, and respect for civil and human rights. Adherence to civil rights laws and human rights principles must be examined in consideration of AI-adoption where rights could be violated.

5. Privacy and Security: AI-enabled tools should be designed in ways that maximize privacy and security of persons and personal data.

6. Human Values: AI-enabled tools should be developed and used in ways that support the ideals of human values, such as human agency and dignity, and respect for civil and human rights. Adherence to civil rights laws and human rights principles must be examined in consideration of AI-adoption where rights could be violated.

7. Shared Benefit and Prosperity: AI-enabled tools should be inclusive and promote equitable benefits (e.g., social, economic, environmental) for all.

8. Accountability: The University of California should be held accountable for its development and use of AI systems in service provision in line with the above principles.

University of California Presidential Working Group on AI
Recommendations

1. Develop a vetting process for AI-enabled tools that affect students.
2. Put in place appropriate safeguards to mitigate risks of algorithmic bias.
3. Educate stakeholders on the risks of AI-enabled tools that affect students.
4. Develop a database that inventories and tracks use of AI-enabled tools used in the “student experience” activities outlined in this section (e.g., in exam proctoring and grading) and put in place transparency procedures.
5. Implement appropriate privacy-preserving methods in the development and use of AI-enabled tools for students.

Conclusions

• What works best?
  • Technological solutions (XAI, Transparent AI)
  • Governance (as in UC)
  • Legal frameworks (e.g. EU AI Act)