

**MUNI**  
**FI**

# **Umělá inteligence a akademická integrita**

Mgr. Tomáš Foltýnek, Ph.D.

[foltynek@fi.muni.cz](mailto:foltynek@fi.muni.cz)

# AI a akademická integrita

- Etika umělé inteligence
  - Co by technologie měla / neměla dělat
- **Důsledky** umělé inteligence na akademickou integritu
  - Automatizované generátory obsahu
  - Důsledky pro vzdělávání
- Umělá inteligence **k zajištění** akademické integrity
  - Detekce plagiátorství
  - Detekce podvodného jednání obecně

# Etika umělé inteligence

- Technické hledisko → Co systém **může** dělat?
- Ethické hledisko → Co by systém **měl** / **neměl** dělat?
  - Přínos pro lidskou společnost
- Někdy nemusí být jasné co je přínosné a co ne
- Někdy je **konflikt** mezi etickými a ekonomickými **hodnotami**

# Etické otázky velkých textových modelů

- Timnit Gebru, bývalá ředitelka Google AI Ethics
- Článek nebyl nikdy publikován, Gebru byla donucena opustit Google
  
- Učení – spotřeba energie / uhlíková stopa
- Učení ze zkreslených dat na internetu
  - Obsah – rasismus, sexismus, násilí (AI vnímá jako běžné)
  - Další znevýhodňování již znevýhodněných skupin
  - Příliš velká data je nemožné auditovat

# Nedávný příklad: Galactica

- Zveřejněna 15. 11. 2022 společností Meta AI
- Generativní jazykový model na pomoc vědcům
  - Naučen na 48 milionech vědeckých článků, učebnic, přednášek,...
- Problémy: Chybný nebo nepřesný, ale přesvědčivý výstup
  - Riziko: Výstupy naruší vědeckou pravdu
  - Navíc k paper mills, predátorským časopisům,...
- Přínosy pro poctivé vědce nejasné
- Stažena po třech dnech

# Důsledky AI pro akademickou integritu



# SClgen

- Scientific paper generator
- Vyvinut v roce 2005 na MIT
- Výstupy **vypadaly** jako vědecký článek
  - Struktura, forma, citace, diagramy,...
- Text byl **naprostý nesmysl**
- 2005: Článek vygenerovaný SClgenem byl přijat na IEEE World Multiconference on Systemics, Cybernetics and Informatics
- Od té doby byl široce využíván čínskými akademiky
- 2013: Springer a IEEE stáhli 122 článků vygenerovaných SClgenem

# GPT-3

- Generative Pre-trained Transformer model, 3. generace
- Vyvinut společností OpenAI v roce 2020
- Jazykový model naučený na kontextu slov
- Učil se z webových stránek, Wikipedie, knih
- Natrénován pro predikci následujících slov
- Na základě krátkého “semínka” vygeneruje libovolně dlouhý výstup
  - Je možné upřesnit žánr, tón (přesvědčivý, přátelský, odborný)
  - Náhodnost – shoda s existujícími zdroji je nepravděpodobná
  - Nerozlišitelný od textu napsaného člověkem



# Text vygenerovaný umělou inteligencí

## Můj vstup:

Implications of the artificial intelligence text generators for university education

## Výstup umělé inteligence (<https://beta.openai.com/playground>):

The implications of the artificial intelligence text generators for university education are far-reaching. The technology has the potential to revolutionize the way in which students learn and how information is disseminated.

One of the most significant implications is that the technology can be used to create personalized learning experiences for students. The text generators can be used to create customized content that is tailored to the individual needs of each student. This would allow for a more efficient and effective learning experience, as students would be able to focus on the material that is most relevant to them.

Another implication is that the technology can be used to create more engaging and interactive learning experiences. The text generators can be used to create texts that are more interactive and allow for more user input. This would create a more immersive and engaging learning experience for students.

Finally, the text generators can be used to create texts that are more easily understandable. The technology can be used to create texts that are easier to read and comprehend. This would allow for a more effective and efficient learning experience, as students would be able to understand the material more easily.

# Automaticky generované texty

Dnes již není potřeba, aby studenti psali eseje. Umělá inteligence je napíše za ně a systémy pro detekci plagiátorství nic neodhalí.

*Studenti si nyní mohou koupit eseje online a získat vysokou známku, aniž by museli vynaložit jakékoli úsilí. Umělá inteligence ji totiž napíše za ně na základě toho, co potřebují, a pak ji vydává za jejich vlastní práci. Jediným problémem je, že studenti tyto služby využívat nesmějí. Musejí tedy používat své vlastní psací dovednosti, což může být pro mnoho studentů obtížný úkol. Budoucnost vzdělávání je ohrožena. S nástupem umělé inteligence se studenti už nebudou muset nic učit. Všechnu práci za ně udělají stroje. To je vážný problém, protože to ze studentů udělá lenochy a hlupáky. Vůbec nic se nenaučí, ale přesto budou mít dobré známky. To by mohlo ovlivnit budoucnost vzdělávání a připravit mnoho učitelů o práci.*

(Vygenerováno systémem copy.ai)

# Umění generované AI

Generated by  
<https://hotpot.ai/art-maker>

## Conference presentation



## Academic integrity



# Umění generované AI

Generated by  
<https://hotpot.ai/art-maker>

**Artificial intelligence**



**Future university education**



# Chat GPT

- Následník GPT-3 a Instruct-GPT
- Vytrénován k interakci s uživatelem
  - A k plnění jeho instrukcí
- Dokáže řešit studentské úkoly
  - Psaní esejů
  - Programovací úlohy
  - Matematické rovnice a jejich soustavy

# Otázky a obavy

- Kdo je autor vygenerovaného obsahu?
  - Jek jej citovat?
- Jak hodnotit studentskou práci?
  - A to i když učitel přesně ví, jak byla vytvořena
- Kde je hranice dovolených nástrojů?
  - Google Translate, DeepL
  - AI na korektury a stylistické vylepšování (Writefull, Grammarly)
  - Je v pořádku, že je využívají akademici? A studenti?

**Ohrožuje**  
**umělá inteligence**  
**akademickou integritu?**

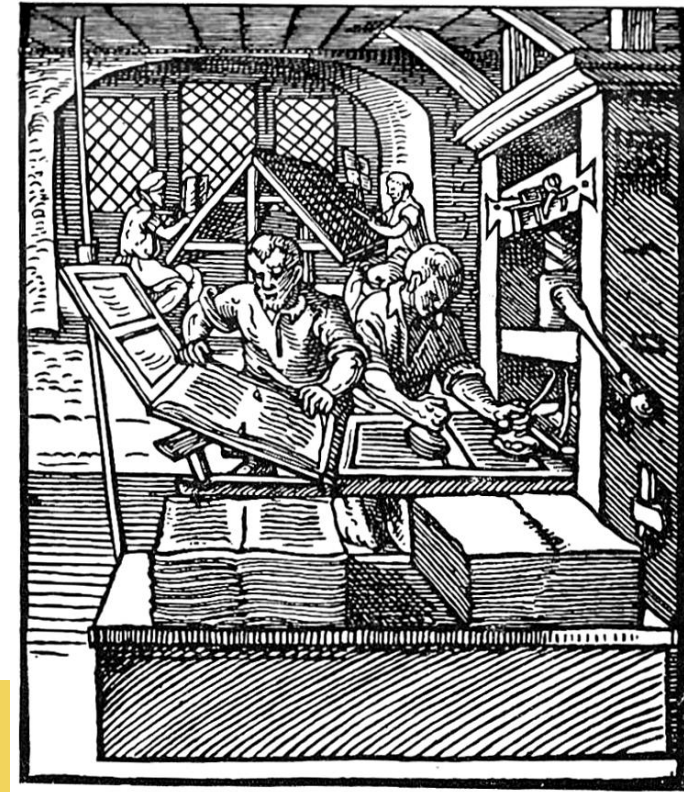
# Příklady technologií, které proměnily vzdělávání





# Knihťisk

- Před: Univerzity měly monopol na vzdělávací materiály
  - Přednáška = předčítání
- Po: Studenti si mohou materiály sami číst
  - Předčítání na přednášce nedává smysl
- Jiné aktivity ve výuce
  - Vysvětlování, diskuse,...



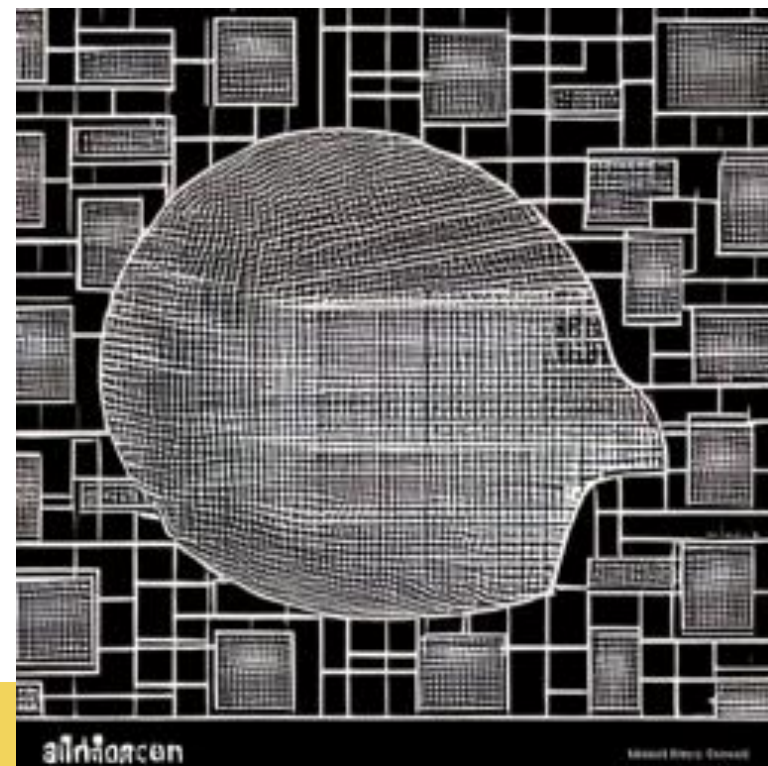
# Kalkulačka

- Vykonává nudné aritmetické operace
- Zakázána při výuce základních aritmetických operací
- Užitečný nástroj ve vyšší matematice



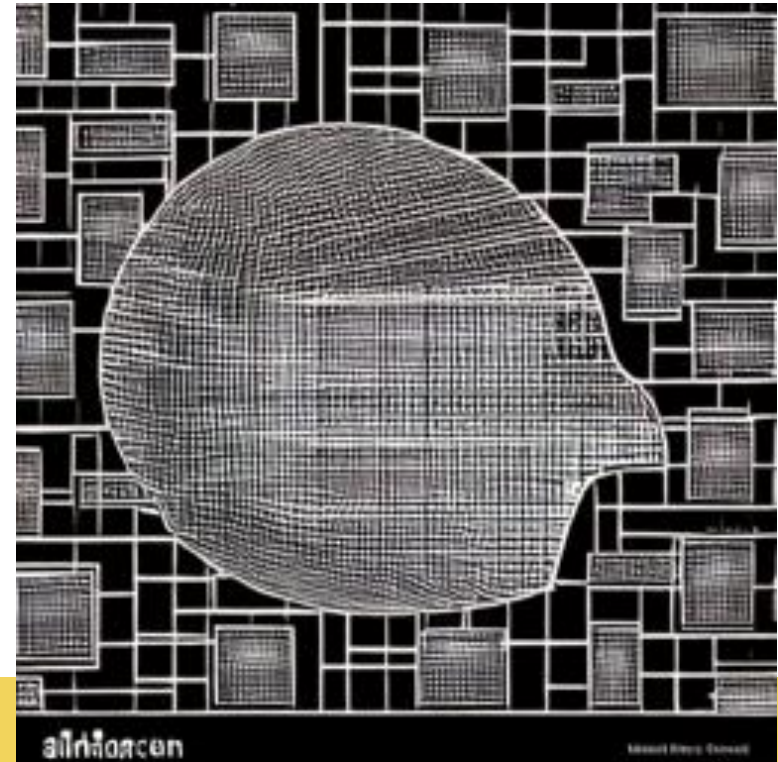
# Automatizované generátory obsahu

- Jen další nástroj, který posune vzdělávání kupředu
- Jakmile studenti získají potřebné dovednosti
  - AI jim může pomoci
  - AI může obohatit výuku



# Důležité jsou výukové cíle

- V určitých fázích učení jsou určité nástroje zakázané
  - Výuka psaní – není povolena tiskárna
  - Výuka základní matematiky – není povolena kalkulačka
- Později se žáci/studenti naučí tyto nástroje využívat
  - Šetří čas
  - Umožní vytvářet lepší výstupy



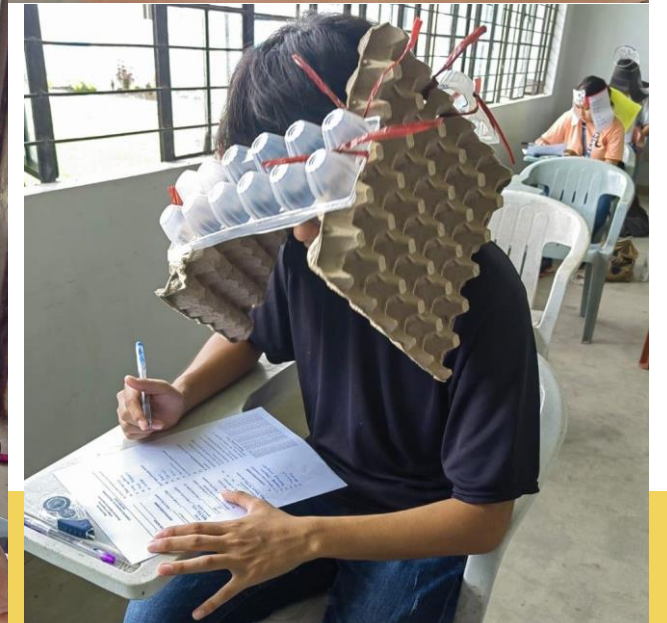
**Ohrožuje**  
**umělá inteligence**  
**akademickou integritu?**

**Můj názor: Ne,**  
**ohrožuje pouze nesmyslné úkoly**

# Zakázané nástroje

- Jak zajistíme, že je studenti nevyužívají?
- Jak zajistíme, že žáci nepoužívají kalkulačku při procvičování násobilků?
- Jak zajistíme, že studenti nepoužívají generátory textu při výuce akademického psaní?

# Philippines: Student 'anti-cheating' exam hats go viral



<https://www.bbc.com/news/world-asia-63363473>

# AI k zajištění akademické integrity





# Proctoring Software

- Hlídnání studentů u online zkoušení
  - Blokuje nedovolené aplikace, snímá obrazovku
  - Rozpoznávání obličeje a podezřelého chování
  - Detekce dalších osob v místnosti
- Problémy
  - Invazivnost, technologické požadavky
  - Snadné obcházení
  - Limity rozpoznávání obličejů (častěji podezřívá ženy a lidi tmavší pleti)
  - Právní otázky
  - Narušení důvěry (základní hodnota akademické integrity)



# Detekce plagiátorství

- Extrinsic plagiarism detection
  - Nalézt potenciální zdroj plagiátorství
  - Hledání podobností mezi dokumenty
- Intrinsic plagiarism detection
  - Nalézt místa, kde se mění styl
  - Hledání rozdílů v rámci jednoho dokumentu
- Současné antiplagiátorské systémy
  - Spolehlivě vyhledávají textovou shodu
- Aktuální výzvy
  - Detekce parafrází, překladu
  - Detekce obsahu vytvořeného někým jiným (něčím jiným)

# Každá mince má dvě strany



- Obava, že kalkulačky zničí výuku matematiky
  - Nezničily. Na 2. stupni ZŠ a na SŠ jsou běžnou součástí výuky
- Obava, že internet usnadní plagiátorství
  - Usnadnil. Ale usnadnil i jeho detekci
  - Studenti mají k dispozici stejné zdroje jako antiplagiátorské systémy
- Obava, že s GPT končí závěrečné práce
  - Nekončí. Jen je třeba se tolik nezaměřovat na text, ale na to, co studenti skutečně udělali
  - Úkoly budou bližší reálné praxi, více persnalizované

# Jak mohou učitelé využít AI

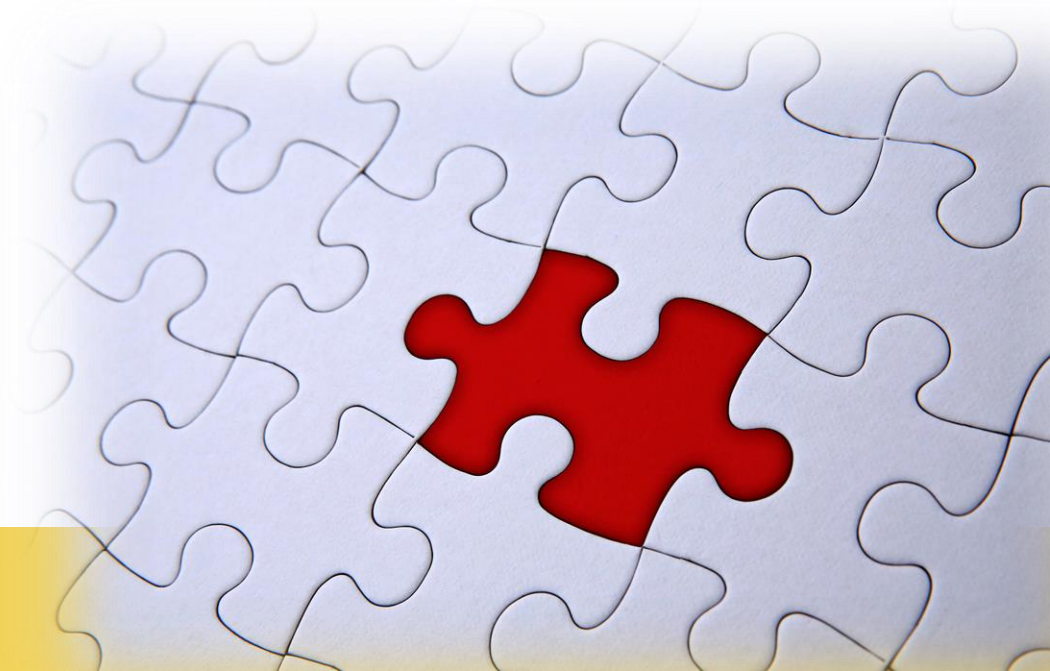
- Hledání inspirace na úkoly pro studenty
  - Není potřeba dlouze "googlit", stačí se zeptat ChatGPT
- Variace zadání, personalizované úkoly
- Využití ve výuce
  - Ať studenti zadejí úkol Chat GPT a zkontrolují, jestli jej vyřešil správně
  - Kde udělal Chat GPT chybu a proč?
- Mají studenti napsat esej?
  - Povolte využití Chat GPT, ale ať jej studenti přiznají
  - Upozorněte je na rizika (nepravdivé výstupy)

# Využití AI ve vědě

- Vědci využívají mnoho nástrojů
  - Některé nástroje využívají umělou inteligenci
- Které nástroje je třeba explicitně uvést?
  - Literární rešerše (Google Scholar)
  - Sběr dat
  - Zpracování dat (statistický software)
  - Nástroj pro psaní testu (Word, Overleaf)
  - Databáze častých slovních spojení (phrasebank)
  - Nástroj pro kreslení obrázků, diagramů, grafů
  - Korektury (Grammarly)
- Ty které významně ovlivnily myšlenku a obsah
  - Obecné kritérium, které bude třeba upřesňovat

# Závěr

- Pouze detekce nestačí
  - Nástroje nebudou nikdy dokonalé
  - AI může pomoci, ale nezastaví podvádění
- Academická integrita má mnoho složek
  - Motivace, autentické hodnocení, hlídání a odhalování, předpisy, procesy, institucionální kultura
- Technologie a pedagogika musí spolupracovat
- Je třeba diskutovat o hranicích etického využití technologií
  - ve vědě i ve vzdělávání



# Děkuji za pozornost

[foltynek@fi.muni.cz](mailto:foltynek@fi.muni.cz)

[tomas.foltynek@academicintegrity.eu](mailto:tomas.foltynek@academicintegrity.eu)

<https://twitter.com/TFoltynek>

