

**MUNI**

# **Etické otázky a výzvy spojené s využitím nástrojů AI v psychologickém výzkumu**

Helena Klimusová

# AI nástroje v psychologickém výzkumu

- nová data pro výzkum
- nové postupy analýzy velkých datových souborů
- modely zpracování přirozeného jazyka
- zapojení konverzačních agentů (chatboti)
- zařízení pro vzdálené monitorování
- generování hypotéz a teorií s pomocí AI
- nástroje pro akademické psaní, vyhledávání zdrojů, rešerše, editaci...

# AI nástroje v psychologickém výzkumu

Bartlett, L. K., Pirrone, A., Javed, N., & Gobet, F. (2023).  
Computational Scientific Discovery in Psychology. *Perspectives on Psychological Science*, 18(1), 178-189.

<https://doi.org/10.1177/17456916221091833>

- poloautomatické vytváření teorií
- modelování procesu rozhodování
- metody pro zpřesnění diagnostických kritérií v klinické psychologii

# Příklady výzkumu s využitím AI

- Stanford University's [Computational Psychology and Well-Being Lab](#) - textová data ze sociálních sítí, měření well-beingu na populační úrovni, možnosti intervence
- Princeton's [Computational Cognitive Science Lab](#) – modelování prvního dojmu na AI generovaných fotografiích obličejů
- Detection and Computational Analysis of Psychological Signals – analýza řečového projevu, mimiky, gest u válečných veteránů, vyhledávání těch, kteří mohou potřebovat psychologickou pomoc

# Etické otázky a výzvy

- otázky spojené s autorstvím a duševním vlastnictvím
- transparentnost, vysvětlitelnost, interpretovatelnost
- důvěrnost informací, ochrana dat, soukromí
- informovaný souhlas
- předpojatost, diskriminace
- odpovědnost

# Autorství /plagiátorství

- úzce spojeno s obecnou otázkou integrity výzkumu
- transparentnost, jasné odlišení vlastního přínosu autora
- riziko nezáměrného plagiátorství
- časopisy postupně začleňují pravidla využití AI do pokynů pro autory, podobně univerzity směrem ke studentům i zaměstnancům
- **APA zásady**
  - uvést použití AI (a podrobně popsat v části metoda)
  - AI nemůže být autor
  - ale musí být citována (podle šablony pro software)
  - jako přílohu dodat celý výstup AI
  - pozor na ochranu dat (zákaz použití generativní AI při recenzích)

# Transparentnost a vysvětlitelnost

- nejvíce diskutované výzvy
- dokumentace souborů dat a způsob zpracování
- schopnost vysvětlit technické procesy a rozhodování
- poskytování informací uživatelům o možnostech a omezeních
- důležité zvláště u „high stakes“ rozhodování
- nejmodernější modely často „černé skříňky“, problém s informovaným souhlasem

# Důvěrnost informací, soukromí

- nástroje AI mohou generovat velké množství osobních údajů
- údaje shromážděné z mnoha zdrojů zvyšují riziko identifikace osoby a tedy porušení soukromí a důvěrnosti



# Informovaný souhlas

- konflikt mezi požadavky na informovanost souhlasu (a z ní plynoucí vědomé a odpovědné rozhodnutí účastníka) a nízkou transparentností/vysvětlitelností nástrojů AI
- je třeba zohlednit již při **konstrukci nástrojů AI**
- některé minimálně rizikové studie souhlas nevyžadují, účastníci si ani nemusí být vědomi použití AI a rozsahu dat
- otázka „**náhodného nálezu**“ – s využitím AI pravděpodobnější, ale problematické opodstatnění

# Předpojatost, diskriminace

- algoritmy AI mohou produkovat předpojaté a diskriminační výsledky
- jazyk spjat se sociokulturními charakteristikami
- skupiny nedostatečně zastoupené při trénování algoritmu ohroženy méně přesným hodnocením
- dopad nejen na účastníky, ale celkově na výzkum

# Odpovědnost

- nutnost právního rámce
- nejasná odpovědnost jednotlivých aktérů (vývojáři, výzkumníci, klinici, prodejci aplikací...)
- k nejvyšší odpovědnosti vedou nástroje s dostatkem důkazů o validitě výstupů, vysokou vysvětlitelností, a systémy UI s otevřeným zdrojovým kódem

# Etické otázky využití nástrojů AI v terapii

- množství aplikací využívajících AI k podpůrné léčbě, monitorování stavu u psychických poruch atd.
- rizika: předpojatost, poškození uživatelů, nerovný přístup k aplikacím, narušení soukromí uživatelů
- chybějící vztah mezi terapeutem a klientem

# Regulace použití nástrojů AI

- směrnice (právní nástroje) – např. EU, WHO
- etické kodexy psychologických společností
- stanoviska a směrnice výzkumných institucí
- pokyny pro autory v odborných časopisech
  
- postupný vývoj speciálních etických kodexů/směrnic pro AI v psychologii
  - Human-in-the-Loop
  - zvláštní zřetel u “černých skříněk