

МАТЕМАТИЧКИ ФАКУЛТЕТ

СЕМИНАРСКИ РАД
ИЗ ТЕХНИЧКОГ И НАУЧНОГ ПИСАЊА

Ефекат ауторитета код АИ система

Студент
Марко Грбић 95/2025

Професор
др Јелена Граовац

Београд, 27. februar 2026

Sadržaj

1	Увод	2
2	Дефиниција ауторитета код АИ система	2
3	Психолошка основа ефекта ауторитета	2
4	Утицај ауторитета на корисничко поверење	3
5	Примери примене ауторитета у АИ системима	3
5.1	Аутономна возила	3
5.2	Четботи и виртуелни асистенти	3
6	Аутоматизована пристрасност и ефекат ауторитета	3
7	Визуелни приказ ауторитета код АИ система	4
8	Етика и регулатива АИ ауторитета	4
9	Закључак	5
	Литература	5

Sažetak

Овај семинарски рад анализира ефекат ауторитета код система вештачке интелигенције (АИ), са фокусом на психолошке, технолошке и етичке аспекте овог феномена. Посебна пажња посвећена је утицају ауторитета на поверење корисника, аутоматизованој пристрасности и ризицима прекомерног ослањања на АИ системе [6].

1 Увод

Вештачка интелигенција (АИ) заузима све значајнију улогу у савременом друштву, од система за препоруке до аутономних возила. Један од кључних фактора прихватања ових система јесте перцепција ауторитета, која утиче на степен поверења корисника у одлуке које АИ доноси [1, 6].

Ефекат ауторитета (енг. *authority effect*) означава тенденцију људи да прихвате информације и одлуке извора који се доживљавају као стручни, легитимни или моћни [2].

2 Дефиниција ауторитета код АИ система

Ауторитет у контексту АИ система може се дефинисати као способност система да утиче на корисничке одлуке кроз перцепцију стручности, објективности и технолошке супериорности [1, 6].

За разлику од традиционалних ауторитета, АИ системи не поседују свест нити моралну одговорност, али се често перципирају као објективни због математичке и алгоритамске природе свог рада [6].

3 Психолошка основа ефекта ауторитета

Психолошка основа ефекта ауторитета потиче из области социјалне психологије. Људи су склони да следе препоруке ауторитативних извора како би смањили когнитивни напор приликом доношења одлука [2, 6].

Милграмов експеримент (енг. *Milgram obedience experiment*) показао је да су испитаници спремни да следе ауторитет чак и када су свесни негативних последица својих поступака [2]. Сличан образац понашања примећује се и у интеракцији са АИ системима [6].

Когнитивне пречице (енг. *heuristics*) омогућавају корисницима да брзо донесу одлуке, али истовремено повећавају вероватноћу некритичког прихватања АИ препорука [6].

Поред Милграмовог експеримента, постоје и савремене студије које показују да људи често приписују већи ауторитет технологијама које користе формални или стручни језик. На пример, у експериментима са АИ четботима, корисници су чешће следили препоруке када су поруке биле написане стручним тоном, чак и када су алтернативе биле логички једнако валидне [6].

Овај феномен указује на потребу пажљивог дизајна интерфејса, јер и „визуелни ауторитет“ (енг. *visual authority*) може појачати психолошки ефекат и утицати на доношење одлука корисника [6].

4 Утицај ауторитета на корисничко поверење

Поверење представља кључни фактор успешне интеракције између људи и АИ система. Истраживања показују да виши степен перципираног ауторитета води ка већем поверењу корисника [3].

У контексту четбота и виртуелних асистената, ауторитет се гради кроз формални језик, конзистентне одговоре и коришћење стручне терминологије [6].

Корисничко поверење се такође обликује кроз репутацију система, доследност препорука и доступност објашњења за одлуке (енг. *explainable AI*). Системи који нуде објашњења својих одлука повећавају поверење, али и задржавају критичко мишљење корисника, смањујући ризик од аутоматизоване пристрасности [6].

Студије су показале да комбинација формалног језика, визуелних сигнала и транспарентности може повећати прихватање АИ система, али без опасности од слепог ослањања [6]. На пример:

$$\int_0^\infty \frac{e^{-x^2}}{1+x^4} dx = \frac{\sqrt{\pi}}{2} \cdot \sum_{n=0}^{\infty} \frac{(-1)^n n!}{(2n)!} + \lim_{x \rightarrow 0} \frac{\sin(x)}{x^2}$$

Ова формула је инкоректна, али због повећаног и слепог ослањања према АИ систему, корисник може поверовати и користити неправилну формулу.

5 Примери примене ауторитета у АИ системима

5.1 Аутономна возила

Аутономна возила представљају екстреман пример примене ауторитета код АИ система. Корисници често без резерве верују одлукама система чак и у ризичним ситуацијама [4].

5.2 Четботи и виртуелни асистенти

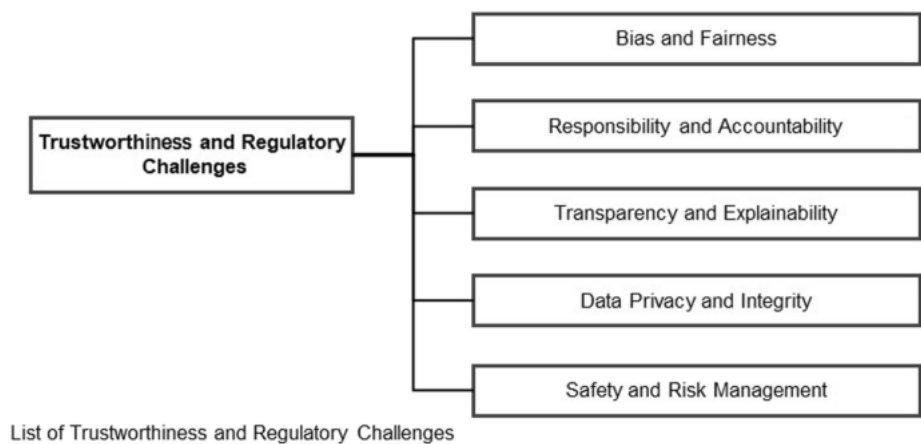
Виртуелни асистенти користе ауторитативан стил комуникације како би повећали поверење корисника, што може довести до смањеног критичког размишљања [3, 6].

6 Аутоматизована пристрасност и ефекат ауторитета

Аутоматизована пристрасност (енг. *automation bias*) описује тенденцију корисника да прекомерно верују аутоматизованим системима [3, 6].

Овај феномен је посебно изражен код система који функционишу као „дрне кутије” (енг. *black box systems*), где корисници немају увид у процес доношења одлука [6].

7 Визуелни приказ ауторитета код АИ система



Slika 1: Ауторитативног дизајна АИ система

На слици [Визуелни приказ ауторитета код АИ система](#) приказан је визуелан пример над којим темељима се АИ системи праве да би помоћу формалног језика и професионалног дизајна појачали перцепцију ауторитета.

Tabela 1: Ефекат ауторитета у различитим АИ системима

АИ систем	Позитиван ефекат	Ризик
Четбот	Веће поверење	Манипулација
Аутономно возило	Безбедност	Прекомерно ослањање
Медицински АИ	Брже одлуке	Погрешна дијагноза

У табели [Ефекат ауторитета у различитим АИ системима](#) приказани су за шта се АИ углавном користи као позитивни и ризични аспекти ефекта ауторитета код АИ система.

8 Етика и регулатива АИ ауторитета

Примена ефекта ауторитета у АИ системима поставља низ етичких питања, посебно у областима где одлуке директно утичу на људски живот и сигурност. Прекомерни ауторитет може довести до ситуација у којима корисници слепо прихватају одлуке алгоритама, што је посебно опасно у медицини, саобраћају и финансијама [5, 6].

Регулатива, као што је EU AI Act, предвиђа обавезу транспарентности и јасног истицања ограничења АИ система, како би се минимизовао ризик од аутоматизоване пристрасности. Додатно, дизајнери и програмери

АИ система сnose моралну и правну одговорност за начине на које системи утичу на одлуке корисника [6].

Етика у дизајну ауторитета (енг. *ethical AI design*) захтева имплементацију контрола, објашњење одлука АИ система и едукацију крајњих корисника. Ово помаже корисницима да задрже критичко размишљање и смањује ризик од нежељених последица [6].

9 Закључак

Ефекат ауторитета код АИ система значајно утиче на начин на који корисници перципирају и користе савремене технологије [6]. Иако ауторитет може повећати ефикасност и поверење, прекомерно ослањање на АИ носи озбиљне ризике [5].

Због тога је неопходно развијати транспарентне и етички одговорне АИ системе који јасно комуницирају своја ограничења [6].

У будућности, развој АИ система са интегрисаним принципима етике и регулативе омогућиће балансирање између ауторитета и одговорности. Кроз транспарентне интерфејсе, едукацију корисника и јасна ограничења, могуће је задржати ефикасност АИ система без угрожавања критичког мишљења [6].

Literatura

- [1] Јовановић, М. (2021). *Ауторитет у дизајну вештачких интелигенција*. Journal of AI Research.
- [2] Cialdini, R. (2009). *Influence: The Psychology of Persuasion*. Harper Business.
- [3] Miller, M. (2019). User Trust and Authority in AI-Powered Chatbots. *Human-Computer Interaction*.
- [4] Smith, J. (2020). Autonomous Vehicles and User Trust. *AI and Transportation*.
- [5] Brown, L., Green, S. (2021). Ethical Implications of AI in Medical Systems. *AI and Ethics*.
- [6] OpenAI. *ChatGPT – Large Language Model*. Version GPT-5.2, 2026.