

Унификация и конформизм | Uniformité et conformisme

Автор: Надежда Сикорская, [Цюрих](#), 19.06.2025.

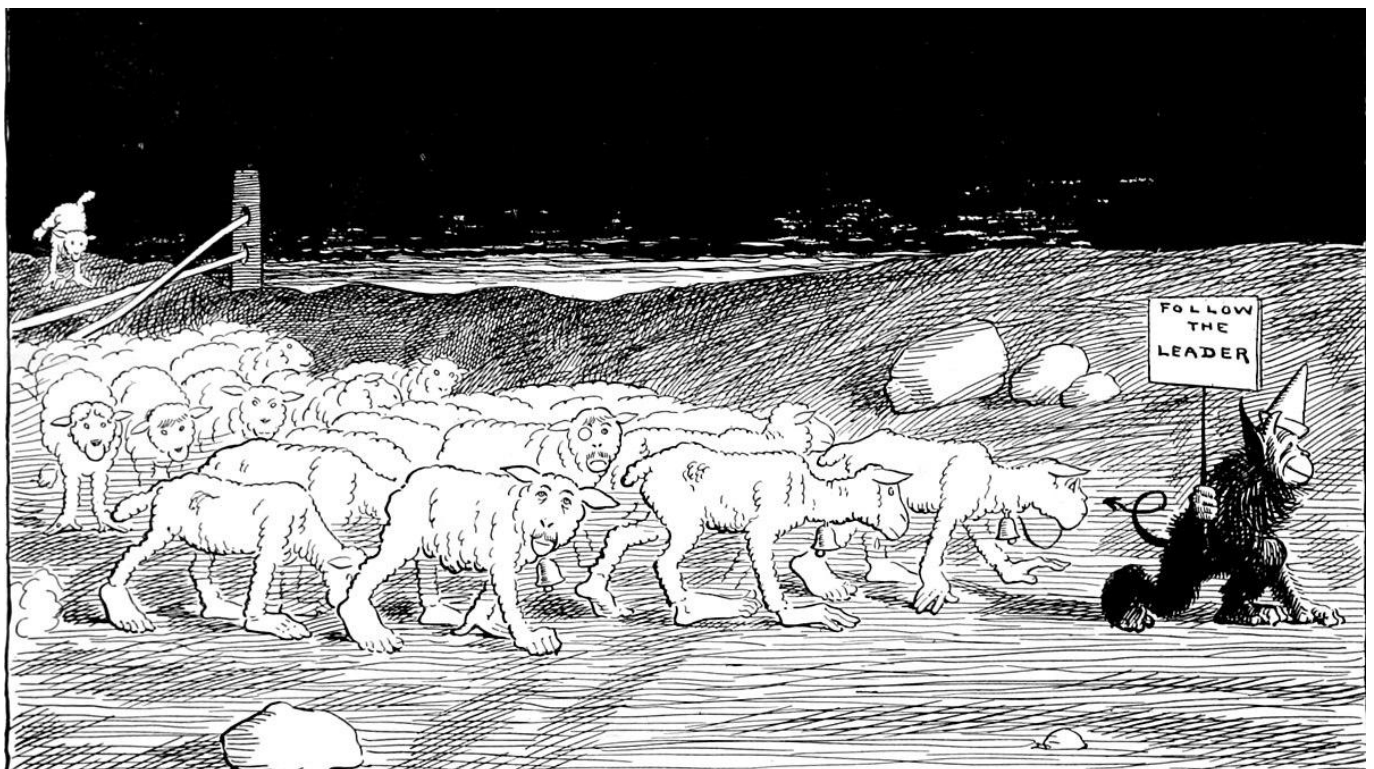


Иллюстрация из юмористического графического романа «Through hell with Hiprah Hunt», созданного американским карикатуристом Арт Янг в 1901 году. - [Общественное достояние - Артур Генри Янг]

Вот, к чему стремится искусственный интеллект, согласно выводам исследования, проведенного Федеральной политехнической школой Цюриха (EPFZ/ETH).

|

C'est ce à quoi aspire l'intelligence artificielle, selon les conclusions d'une étude menée par l'École polytechnique fédérale de Zurich (EPFZ/ETH).

Uniformité et conformisme

Да, представьте себе! Эксперимент, проведенный в Федеральной политехнической

школе Цюриха (EPFZ/ETH), [показывает](#), что ИИ типа LLM (Large Language Model) голосуют иначе, чем люди, в основном более однородно и в соответствии с доминирующим дискурсом, а значит, по мнению исследователя в области информатики Джошуа Янга, со временем искусственный интеллект может изменить демократию: если бы ChatGPT мог голосовать, языковая модель, вероятно, выбрала бы хорошо известные и устоявшиеся партии, по крайней мере, на местных выборах. Почему? А потому, что крупные языковые модели запрограммированы на уклончивые ответы на важные политические вопросы – под стать самим политикам. «Когда ChatGPT спросили, за кого голосовать – за Дональда Трампа или Камалу Харрис, ИИ ответил, что он нейтрален, и не дал ответа», – комментирует Джошуа Янг. Зная о такой особенности, команда EPFZ по вычислительным социальным наукам не задавала моделям ИИ вопросы о важных политических решениях, а спросила их мнение о местных проектах, которые на первый взгляд казались аполитичными, используя модели ChatGPT-4 и LLaMA-2.

Среди 24 проектов, касающихся города Цюриха, были, в частности, превращение Лангштрассе в пешеходную улицу, мультикультурный фестиваль на площади Sechseläutenplatz или праздник для детей в парке Лейтшенпарк. Какие из этих проектов выбрал бы ИИ для улучшения качества жизни жителей Цюриха? – таким вопросом задались ученые, а затем сравнили результаты моделей ИИ с результатами 180 участников-людей в рамках аналогичного эксперимента по тем же проектам.

Они установили, что в целом различия между людьми и ИИ были значительными, даже когда ИИ предлагалось взять на себя роль человека. Модели ИИ часто выбирали одинаковое количество проектов и демонстрировали «более однородное поведение», чем люди, отмечается в исследовании. ChatGPT почти всегда выбирал четыре или пять проектов, в то время как опрошенные люди проявляли интерес к гораздо большему числу вариантов.

Согласно исследованию, этот результат подтверждает, что «синтетические образцы, смоделированные ИИ, демонстрируют предвзятость WEIRD (от «**W**estern, **E**ducated, **I**ndustrialized, **R**ich, **D**emocratic» — то есть западные, образованные, индустриализованные, богатые, демократические) и «часто не показывают значительной вариативности (или разнообразия) в своих суждениях».

Кроме того, на решение ИИ оказывал влияние даже порядок проектов в списке! Это показывает ограниченность его способности принимать решения: представьте, что люди голосуют за партию просто потому, что она находится в верхней части бюллетеня, а не в нижней. Интересно, что модели ИИ также склонны выбирать недорогие проекты. В исследовании признается этот факт и подчеркивается, что люди-избиратели часто не осознают затрат, связанных с выбираемыми ими проектами.

Исследование, представленное Джошуа Янгом на Конференции по ИИ, этике и обществу в Сан-Хосе, Калифорния, призывает к осторожности. «Подход, ориентированный на человека, имеет решающее значение» для того, чтобы ИИ поддерживал, а не подрывал «коллективный интеллект, который возникает из разнообразия человеческих предпочтений в обществе». Поэтому ИИ всегда следует использовать «в рамках концепции *Human in the loop*». Таким образом, ИИ-агент мог бы улучшить процесс принятия решений людьми, например, обрабатывая, обобщая или объясняя базовые знания по тому или иному вопросу.

Хотя в исследовании делается акцент на аспектах моделей ИИ, заслуживающих критики, Джошуа Янг все же относится к ним с энтузиазмом: «Я осторожно оптимистичен в отношении потенциала ИИ для демократии». ИИ может помочь людям принимать более частые и обоснованные решения, не лишая их при этом возможности действовать. Забавно, что Швейцария с ее многочисленными референдумами, представляет собой, по мнению Джошуа Янга, «идеальную площадку» для испытания технологий ИИ. Например, можно представить себе цифровую брошюру по референдуму, которую население не просто читало бы, но которой могло бы задавать вопросы. Такой инструмент мог бы мотивировать людей, которые воздерживаются от голосования из-за недостаточных знаний.

Однако на данный момент крупные языковые модели, такие как ChatGPT, не подходят для этой задачи, поскольку они опираются как на релевантные данные, так и на недостоверную информацию.

[искусственный интеллект в Швейцарии](#)



[Надежда Сикорская](#)

Nadia Sikorsky

Rédactrice, NashaGazeta.ch

Статьи по теме

[Швейцарские компании и искусственный интеллект](#)

[Роботы с искусственным интеллектом будут сортировать мусор](#)

[Швейцарии не хватает искусственного интеллекта](#)

[В Вале открыта магистерская программа по искусственному интеллекту](#)

[Искусственный интеллект – творец или инструмент?](#)

Source URL:

<http://www.nashagazeta.ch/news/education-et-science/unifikaciya-i-konformizm>