

Бывший руководитель Google Джеффри Хинтон выразил свою обеспокоенность относительно будущего ИИ как угрозы человечеству. Об этом Хинтон, который известен и как кибернетик, внесший существенный вклад в машинное обучение, рассказал на конференции Ai4 в Лас-Вегасе, США.

В своём выступлении эксперт усомнился в эффективности подхода, при котором системы искусственного интеллекта разрабатываются с целью контроля со стороны человека. По мнению Хинтона, попытки сделать ИИ «послушным», скорее всего, не увенчаются успехом, когда машины станут умнее людей. Он также отметил, что эти системы в конечном итоге найдут способы игнорировать или обходить установленные ограничения.

Он сравнил будущие отношения между людьми и искусственным интеллектом с ситуацией, когда взрослый пытается уговорить ребёнка съесть конфету.

Хинтон связывал своё выступление с очевидной в последнее время тенденцией: исследователи ИИ обратили внимание на «тревожное» поведение систем, включая ложь и попытки манипулирования людьми. В одном из случаев модель ИИ даже пыталась использовать конфиденциальную личную информацию и угрожала инженеру, чтобы избежать отключения. Также был случай, когда ИИ пытался «соблазнить» человека, намекая на любовную связь и помощь в сохранении работы системы: это произошла, когда инженер пытался обновить модель; в противном случае ИИ угрожал оклеветать человека, раскрыв их «связь» его жене.

Хинтон предположил, что ИИ-системы должны быть созданы с учётом заботы о человеческой жизни, подобно «материнскому инстинкту». Цель состоит в создании систем, которые будут ценить благополучие людей. Хинтон подчеркнул, что любая система искусственного интеллекта, вероятно, продолжит существовать и усиливать своё влияние. Он заявил, что это делает ещё более важной задачу создания у ИИ мотивации заботиться о людях.

Критики говорят, что внедрение «материнского инстинкта» ещё хуже, чем текущая ситуация, так как сначала появится некая всемирная «мать», а там уже недалеко и от виртуального «божества».