

# Пресс-релиз

11 мая 2020 г.

## Открытие чёрного ящика ИИ: модель распознавания лица xCos

11 мая Министерство науки и технологий Тайваня (MOST) провело пресс-конференцию, чтобы представить xCos, модуль проверки лица с объяснимостью, который был разработан командой во главе с профессором Уинстоном Сюй (Prof. Winston Hsu) в Объединенном исследовательском центре технологий искусственного интеллекта и перспектив здравоохранения (MOST Joint Research Center for AI Technology and All Vista Healthcare, AINTU). Помимо распознавания, этот модуль может объяснить, почему ИИ считает, что эти два изображения лица являются идентичными и почему нет. Включение объяснимости решений ИИ будет не только продвигать технологии ИИ, но и обеспечит доверие людей.

По словам министра науки и технологий, д-ра Лян-Цзи Чэнь (Dr. Liang-Gee Chen), Министерство науки и технологий Тайваня оказывает поддержку четырем исследовательским центрам ИИ для разработки различных областей исследований с 2018 года, после того как в 2017 году была объявлена научная стратегия ИИ. Эти центры, расположенные в Национальном Тайваньском университете (National Taiwan University), Национальном университете Цин Хуа (National Tsing Hua University), Национальный университет Цзяо Тун (National Chiao Tung University) и Национальный университет Чэн Гун (National Cheng Kung University), специализируются на технологиях искусственного интеллекта, биомедицинских технологиях, интеллектуальном производстве, прикладных исследованиях искусственного интеллекта, гуманитарных и социальных науках. Команда профессора Сюй из НТУ (National Taiwan University) сосредоточена на технологии искусственного интеллекта; их достижение - модуль xCos, который можно подключить к любым существующим моделям глубокой проверки лиц. Тем временем они готовятся распространить объясняемый ИИ (XAI) на различные области, такие как энергетика, медицина и обрабатывающая промышленность.

### Лучшее в мире распознавание лиц

В 2011 году команда профессора Сюй разработала первую поисковую систему человеческого лица на мобильных устройствах. С тех пор они

продолжали решать проблемы в нескольких областях, включая распознавание лиц разных возрастов и замаскированных лиц. В 2018 году они заняли первое место в категории «распознавание замаскированных лиц» (Disguised Face Recognition) на конференции по компьютерному зрению и распознаванию образов (The Conference on Computer Vision and Pattern Recognition, CVPR).

Профессор Сюй рассказал, что в течение последних нескольких лет они оказывали помощь многим компаниям, занимающимся разработкой программного обеспечения, в разработке технологий распознавания лиц в рамках сотрудничества между академическими кругами и промышленностью. Обнаружение, что некоторые результаты проверки ИИ являются нелогичными, вдохновило их на то, чтобы потратить больше времени на создание xCos, чтобы изучить обоснование решений ИИ. Эта модель может предоставить как количественные, так и качественные причины для объяснения того, принадлежат ли два изображения лица одному и тому же человеку или нет. Если два изображения лица рассматриваются моделью как одно и то же лицо, то метод, предложенный группой, может четко показать, какие области лица являются более репрезентативными, чем другие, путем предоставления локальных значений сходства и внимания. Объясняемый модуль xCos может даже хорошо работать с другими распространенными системами распознавания лиц в нейронной сети, такими как ArcFace, CosFace и т. Д.

### **Демистификация черного ящика ИИ**

Министр Чэнь сказал, что причина, почему технологии постоянно совершенствуются, заключается в том, что люди постоянно спрашивают «почему» и прилагают усилия для поиска решений. На ранней стадии ИИ управлялся системой, основанной на правилах. Было довольно легко проследить, как система приняла свое решение, и это было весьма объяснимо. В настоящее время с увеличением объема данных, глубокие сверточные нейронные сети достигают более высокой точности для задачи проверки лица. Однако люди заметили, что ИИ не может объяснить процесс принятия решений. Учитывая исключительные характеристики ИИ, модели глубокой проверки лиц нуждаются в большей интерпретации, чтобы мы могли доверять результатам, которые они генерируют.

Согласно компании консалтинга и аудита PricewaterhouseCoopers, ИИ обладает потенциалом рыночной стоимости в 15 триллионов долларов, но трудно понять, как алгоритм приходит к своей рекомендации или решению – как работает «объясняемый ИИ». В настоящее время сообщества компьютерного зрения по-прежнему не имеют эффективного метода для понимания механизма работы моделей глубокого обучения. Из-за их врожденных нелинейных структур и сложного процесса принятия решений

(так называемый «черный ящик»), чьи неизвестные причины могут привести к серьезным проблемам безопасности и конфиденциальности. Вышеупомянутые проблемы заставят пользователей чувствовать себя незащищенными в отношении систем, основанных на глубоком обучении, и заставят разработчиков бороться за их улучшение. Таким образом, мы можем себе представить, почему объясняемый ИИ (XAI) рассматривался в рамках проекта Управления перспективных исследовательских проектов Министерства обороны США (Defense Advanced Research Projects Agency, DARPA) на \$ 2 млрд. в 2018 году.

### **Объяснимость и высокая совместимость**

Министр Чэнь добавил, что объясняемый ИИ, безусловно, является важным вопросом в развитии ИИ в мире. С учетом этой тенденции в сентябре прошлого года Министерство науки и технологий обнародовало «Руководство по исследованиям и разработкам в области ИИ», чтобы подчеркнуть «прозрачность и прослеживаемость» и «объяснимость». В феврале Евросоюз также опубликовал БЕЛЫЙ ДОКУМЕНТ об искусственном интеллекте (WHITE PAPER On Artificial Intelligence) - европейский подход к совершенству и доверию, в котором отмечалось, что отсутствие доверия блокирует широкое использование ИИ. Будущее развитие ИИ должно заставить людей доверять ИИ путем укрепления прозрачности и объяснимости.

Профессор Сюй сказал, что xCos может помочь людям понять, как были приняты решения, и помочь разработчикам изучить идеи глубоких нейронных сетей. Он предоставил исходный код на GitHub (<https://github.com/ntubiolin/xcos>) для исследовательского сообщества. Он считает, что благодаря методам объяснимости люди будут более уверены в принятии решений ИИ, а разработчики могут корректировать свои программы для повышения точности.

### **Media Contact**

Winston H. Hsu

Prof. of CSIE, National Taiwan University

The associate editor for IEEE Trans. on Multimedia (TMM), and Trans. on Circuits and Systems for Video Technology (TCSVT)

NVIDIA AI Lab, Founding Director

Co-founder of thingnario

CSIE, National Taiwan University

Tel: 886-2-3366-4888

Email: [whsu@ntu.edu.tw](mailto:whsu@ntu.edu.tw)

Shao-Ping Chiang

Officer, Department of Foresight and Innovation Policies, MOST

Tel: 886-2-2737-7982

E-mail : [spchiang@most.gov.tw](mailto:spchiang@most.gov.tw)