معرفی پنجاهمین وبینار تخصصی انجمن رمز ایران
زمان سخنرانی چهارشنبه ۱۷ دی ماه ۱۴۰۴ ساعت ۱۵

معرفی سخنران:

آقای مهندس علیرضا آقاباقرلو دوره کارشناسی مهندسی برق را از دانشگاه تبریز در سال ۱۳۹۷ و کارشناسی‌ارشد مهندسی برق را از دانشگاه صنعتی شریف در سال ۱۳۹۹ دریافت کرده اند. ایشان در حال حاضر، دوره‌ی دکتری خود را در گروه **COSIC** از دانشکده مهندسی برق دانشگاه کی‌یو لوون **KU Leuven** دنبال می‌کنند. علاقه‌مندی‌های پژوهشی وی شامل بهبود کارایی مدل‌های یادگیری ماشین، تحلیل ویژگی‌های پایدار و ناپایدار در شبکه‌های عصبی عمیق و بررسی کاربرد مدل‌های زبانی است.

عنوان سخنرانی:

تحلیل پایداری مدل‌های یادگیری ماشینی و توسعه ویژگی‌های قابل اعتماد در مدل‌های زبانی

Analysis of the Stability of Machine Learning Models and Extending Reliable Features to Language Models

چکیده سخنرانی:

این ارائه به بررسی نقش هوش مصنوعی در بهبود سامانه‌های دیجیتال و تحلیل نقاط قوت و ضعف مدل‌های یادگیری عمیق می‌پردازد. ابتدا توضیح داده می‌شود که چرا برخی ویژگی‌ها در مدل‌ها واقعاً پایدار و قابل اعتماد هستند، در حالی که برخی دیگر تنها ظاهری قابل اطمینان دارند و در عمل رفتار مطمئنی نشان نمی‌دهند. روش پژوهشی ارائه شده، امکان شناسایی این ویژگی‌ها و ساخت مدل‌هایی با رفتار قابل اعتمادتر را فراهم می‌کند. در بخش بعد، اثر تکرار نمونه‌ها در فرآیند آموزش مدل‌ها بررسی می‌شود. نتایج نشان می‌دهند که این تکرار می‌تواند باعث تغییر تعادل اطلاعات و اثرگذاری بر کیفیت خروجی مدل‌ها شود. همچنین، پژوهش به کاربرد مدل‌های زبانی پرداخته و نشان می‌دهد که این مدل‌ها، با وجود اندازه و پیچیدگی بالا، نیازمند تحلیل دقیق ویژگی‌های پایدار و قابلیت اعتماد برای استفاده در سامانه‌های واقعی هستند. در نهایت، ارائه نکاتی در خصوص آماده‌سازی داده‌ها، انتخاب ویژگی‌ها و روش‌های ارزیابی عملکرد مدل‌ها ارائه می‌کند تا عملکرد نهایی سامانه‌های هوش مصنوعی ارتقاء یابد و تصمیم‌گیری بر اساس آن‌ها مطمئن‌تر شود.