



Contribution ID: 59

Type: **Keynote speech**

## **[PL] Bielik: droga do polskiego modelu językowego (LLM)**

*Friday, 4 April 2025 09:00 (30 minutes)*

W trakcie prezentacji szczegółowo zostanie omówiona historia modeli Bielik, począwszy od rozpoczęcia współpracy między SpeakLeash i ACK Cyfronet AGH, aż po publikację różnych wersji modelu, takich jak Bielik v0.1 i Bielik v2.0. Kolejnym ważnym elementem prezentacji będzie omówienie etapów przygotowania modeli LLM, takich jak wybór modelu i podejścia, tokenizacja, trening bazowy, finetuning instrukcyjny oraz optymalizacja. Zostaną również przedstawione metody poprawy jakości danych, w tym deduplikacja, czyszczenie i anonimizacja, które są kluczowe dla uzyskania wysokiej jakości wyników. Na koniec przedstawione zostaną przypadki zastosowania Bielika. Wśród nich znajdą się odpowiedzi na pytania, w jaki sposób model może być dostosowany do specyficznych potrzeb i jak jego elastyczność i niskie koszty użytkowania czynią go atrakcyjnym wyborem dla różnych zastosowań.

**Presenter:** KINAS, Remigiusz

**Session Classification:** S5