Kaggle | Google

5-Day Gen AI Intensive Course with Google 2025

Whitepaper Companion Podcast (Notes): Solving Domain-Specific Problems Using LLMs: Cybersecurity and Medicine

Introduction to LLMs and Domain Specialization

- Large Language Models (**LLMs**) are now being used to solve complex problems in specialized fields[1].
- Fine-tuning LLMs for specific areas yields significant results and opens new possibilities[1].
- The whitepaper explores the challenges and opportunities of specialized data, technical language, and sensitive use cases in cybersecurity and medicine[1].

Cybersecurity Challenges and SecLM

- Cybersecurity faces challenges such as limited public data, diverse technical concepts, and rapidly changing threats[1].
- Sensitive use cases like malware analysis require specific model development considerations[1].
- **SecLM** is a security-focused language model paired with supporting techniques for threat identification and risk analysis[1].

Pressures in Cybersecurity

- Constant emergence of new and sophisticated attacks[1].
- Operational toil for security teams[1].
- Shortage of skilled professionals[1].

LLMs as AI Assistants in Cybersecurity

- LLMs can act as AI assistants to handle tedious tasks, freeing experts for strategic work[1].
- Examples of LLM applications:
    - Translating natural language into complex query languages[1].
    - Automating investigation and categorization of alerts[1].
    - Generating personalized remediation plans[1].
    - Reverse engineering and understanding malicious software[1].
    - Generating summaries of the threat landscape[1].
    - Providing insights into potential attack pathways and preventative measures[1].
    - Identifying critical areas for security testing and generating secure code snippets[1].

Layered Approach in Cybersecurity

- Existing security tools provide data and context[1].
- A specialized model API, such as SecLM, is in the middle[1].
- Authoritative security intelligence and human expertise underpin the entire system[1].

SecLM as a Central Resource

- SecLM is envisioned as a central resource for security questions[1].
- It aims to be a one-stop shop where users can ask questions in plain language and receive answers based on internal data sources[1].

Standards for SecLM

- Timeliness: Keeping the model up to date with the rapidly changing threat landscape[1].
- Data Sensitivity: Analyzing sensitive user data without risk of exposure[1].
- Deep Security Knowledge: Understanding security concepts and terminology[1].
- Multi-Step Reasoning: Combining different data sources and models to solve problems[1].

Reasons General-Purpose LLMs Fall Short

- Limited publicly available, high-quality security data[1].
- The breadth and depth of security knowledge required[1].
- Sensitive use cases like analyzing malware[1].

Creating Specialized SecLMs: Targeted Training Approach

- Start with a strong general-purpose foundation model with multilingual capabilities[1].
- Pre-training on cybersecurity-specific content (blogs, threat reports, detection rules, IT security textbooks)[1].
- Supervised fine-tuning on tasks that mirror real-world security expert activities:
    - Analyzing potentially malicious groups[1].
    - Explaining command-line instructions[1].
    - Interpreting security event logs[1].
    - Summarizing complex threat reports[1].
    - Generating queries for security management platforms[1].
- Focus on privacy, keeping user-specific data separate[1].

Evaluating Performance of Specialized Models

- For tasks with clear answers, use standard classification metrics[1].
- For open-ended tasks, compare the model's output to expert-provided answers using metrics like Rouge and BERTScore[1].
- Use larger LLMs for automated side-by-side comparisons[1].
- Human evaluators play a crucial role in judging the models' performance[1].

Techniques to Help Models

- **In-context learning:** Adapting to new security platforms by providing examples[1].

- **Parameter Efficient Tuning (PET):** Customizing the model with user-specific data without retraining the entire model[1].
- **Retrieval Augmented Generation (RAG):** Keeping the model up to date with the latest threat intelligence by pulling information from external sources in real-time[1].

Flexible Planning and Reasoning Framework

- Illustrative example: Responding to a security analyst's high-level question about the AP41 threat group[1].
- The SecLM API orchestrates a series of steps:
    - Retrieving information about AP41[1].
    - Extracting key TTPs and IOCs[1].
    - Translating that into a query[1].
    - Running the query against the user's SIEM system[1].
- The sequence of actions can be predefined or generated in real-time[1].
- Automation can save analysts hours of work[1].

SecLM Applications

- Interacting with external security tools using RAG[1].
- Employing specialized models for specific analytical tasks[1].
- Utilizing a form of long-term memory to remember user preferences and context[1].

Ultimate Goal for SecLM

- To become a central platform that transforms how cybersecurity is practiced[1].
- Significantly reduce the daily burden on security professionals[1].

Healthcare and Med-PaLM

- LLMs can transform medical question answering[1].

- **Med-PaLM** is an LLM adapted from Google's PaLM family, focused on improving health outcomes[1].

Potential Uses of GenAI in Healthcare

- Patients can ask questions about their medical history and receive personalized guidance[1].
- AI systems can triage patient messages to the right clinicians[1].
- Revolutionize patient intake[1].
- Provide real-time feedback during consultations[1].
- AI consultant with access to a vast body of medical knowledge[1].

Responsible Innovation in Medicine

- The paper emphasizes the importance of responsible innovation, especially where patient safety is paramount[1].
- Rigorous validation is needed through retrospective analysis and prospective studies[1].

Shift in Scientific Approach

- The vision is to create more human-centered AI systems that can interact with people in a more natural way[1].
- It's about language, empathy, and understanding the human element[1].
- Med-PaLM is presented as a first step towards this vision, starting with question answering[1].

Med-PaLM Progress

- Med-PaLM was the first AI to surpass the passing score on USMLE-style medical license exams[1].
- Med-PaLM 2 achieved expert-level performance on those exams[1].
- It also showed improvements in the quality and depth of its long-form answers[1].

Measuring AI's Medical Knowledge: Evaluation Strategy

- Combine quantitative metrics with qualitative assessments[1].
- Use USMLE-style questions as a benchmark[1].
- Qualitative assessments look at factual correctness, appropriate use of medical knowledge, helpfulness, potential biases, and the potential for harm[1].

Human Evaluations

- Med-PaLM and teams of physicians answer the same medical questions independently[1].
- Responses are given to expert raters who compare them side by side[1].
- The focus is on the substance of the answer[1].

Areas for Improvement

- Models still need to improve, and scoring well on datasets doesn't guarantee real-world performance[1].
- A progression of studies is needed[1].

Task-Specific vs. Broad Domain Models

- Med-PaLM's success shows the value of domain specialization[1].
- Each application needs to be carefully validated and adapted[1].
- The multimodal nature of medicine requires integrating information from images, EHRs, sensor data, and genomics[1].

Applications Beyond Patient Care

- Scientific discovery (identifying genes associated with specific traits)[1].

Med-PaLM as a Suite of Commercially Available Models

- Built on Med-PaLM 2, this will allow healthcare organizations to build their own GenAI solutions[1].

Med-PaLM 2 Training

- Builds upon the base LLM PaLM 2[1].
- Fine-tuned using a lot of medical question answering data[1].
- Uses prompting techniques for multiple-choice questions (few-shot prompting, Chain of Thought prompting)[1].
- Self-consistency helps improve accuracy[1].
- Ensemble refinement is a technique where the model takes into account its own generated explanations[1].

Conclusion

- LLMs show incredible potential in cybersecurity and healthcare[1].
- In cybersecurity, SecLM can automate tasks, address the talent shortage, and revolutionize security practices[1].
- In healthcare, Med-PaLM is tackling the complexity of medical data and improving healthcare delivery[1].
- Collaboration with clinicians and careful real-world evaluation are crucial[1].
- The development of vertical-specific foundation models points to a future where AI is deeply integrated into healthcare[1].