Self-Study AI/ML/LLM Roadmap for Security Folks

This roadmap is coming from a security enthusiast who wanted to improve herself in AI/ML/LLM area to discover its security side.

Small tips before jumping in it:

- You can have the Coursera courses in audit-only mode. You should only know that you can't get a course certificate or grades in audit mode.
    - https://www.coursera.support/s/article/209818613-Enrollment-options?language=en_US
    - https://www.classcentral.com/report/coursera-signup-for-free/
- You can jump to security playgrounds without waiting to finish the training list. No harm in poking around and jumping back. Sometimes, you may not know what you will have exploited. So, never give up, long run but worth it.
- I strongly recommend building some small apps to understand how it works!
- Check MLU, it's made by AWS ❤️ You can find the playlists here: https://www.youtube.com/@machinelearninguniversity1942/playlists

Roadmap:

1. The start point is a good detailed entrance training: Introduction to Machine Learning in Production course from DeepLearning.AI - https://www.coursera.org/learn/introduction-to-machine-learning-in-production/
2. To get much more know-how about neural networks, supervised learning, and regression problems, this course can be a second step for this roadmap. https://www.coursera.org/learn/neural-networks-deep-learning/
3. This course is developed by DeepLearning.AI. At the end of the course, you'll gain foundational knowledge, practical skills, and a functional understanding of how generative AI works. Hopefully, you'll be able to dive into the latest research on Gen AI to understand how companies are creating value with cutting-edge technology- Instruction from expert AWS AI practitioners who actively build and deploy AI in business use cases today. https://www.coursera.org/learn/generative-ai-with-llms/ As another option, this course could be quite useful: https://aws.amazon.com/blogs/machine-learning/new-technical-deep-dive-course-generative-ai-foundations-on-aws/
4. The prompt is everything! You need to learn how you can talk with an LLM! https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/
5. First steps to prompt injection attacks: https://research.nccgroup.com/2022/12/05/exploring-prompt-injection-attacks/
6. Small Playgrounds to keep poking around:

- Gandalf LLM Challenge - https://gandalf.lakera.ai/
  - https://medium.com/the-abcs-of-ai/gandalfs-challenge-mastering-prompt-engineering-for-ai-success-fd777be2aa0b
- Mosscap LLM Challenge - https://grt.lakera.ai/mosscap
- DoubleSpeak Chat - https://doublespeak.chat
  - https://doublespeak.chat/#/handbook

7. More know-how about neural networks and training cycles! Advanced Learning Algorithms - https://www.coursera.org/learn/advanced-learning-algorithms
8. This part may require a bit more math background than the others. But it's a gem to learn about Reinforcement Learning. Unsupervised Learning Recommenders Reinforcement Learning - https://www.coursera.org/learn/unsupervised-learning-recommenders-reinforcement-learning#modules
9. Deep Neural Network - https://www.coursera.org/learn/deep-neural-network#modules
10. NLP Sequence Models - https://www.coursera.org/learn/nlp-sequence-models#modules

References for Security / Essential to Look

1. OWASP Machine Learning Security Top 10 - https://owasp.org/www-project-machine-learning-security-top-10/
2. OWASP Top 10 for LLM - https://owasp.org/www-project-top-10-for-large-language-model-applications/
3. MITRE ATLAS - https://atlas.mitre.org/
4. HuggingFace.co is the ‘GitHub’ of LLMs and machine learning. This is a central point for ML models, datasets, and other content. The transformers library, maintained by HuggingFace, is the cornerstone of the foundation in open-source LLMs. https://huggingface.co/
5. LLM Agency - LlamaIndex (formerly GPT Index) - LangChain, transformers, agents, llamaindex and other similar libraries or custom codes give LLMs agency. This concept is an agency in the truest meaning of the word. These libraries and the concept of agency expose tools and capabilities to an LLM via prompting and then give the model a task. The model is iteratively prompted by making a plan and then drilling down through each step until it gets a final answer. This is where the major new functionality sits and where the new vulnerabilities will be found.

   https://gpt-index.readthedocs.io/en/latest/
6. LLM Agency - Langchain - https://github.com/langchain-ai/langchain
7. LLM Agency - Transformers - https://huggingface.co/docs/transformers/transformers_agents
8. Some repositories that I got good info

   https://github.com/jiep/offensive-ai-compilation

   https://github.com/unica-mlsec/mlsec

https://github.com/Trusted-AI/adversarial-robustness-toolbox/wiki/ART-Attacks

https://github.com/Trusted-AI/adversarial-robustness-toolbox

YouTube Channels to Stay Updated

I've got all those good channel advice from Garrett! (^^,) Thank you so much!

https://www.youtube.com/@matthew_berman

https://www.youtube.com/@mreflow

https://www.youtube.com/@YannicKilcher

https://www.youtube.com/@HuggingFace

https://www.youtube.com/@RobertMilesAI

https://www.youtube.com/@HeatonResearch

https://www.youtube.com/@NicholasRenotte

https://www.youtube.com/@reidhoffman

https://www.youtube.com/@testingai/videos

https://www.youtube.com/@engineerprompt

https://www.youtube.com/@MachineLearningStreetTalk

https://www.youtube.com/@Deeplearningai/videos

https://www.youtube.com/@DrAlanDThompson