

Artificial Intelligence (AI) 101:

What is it? Who's using it? How can I use it?

Presented to Rotary Club of Honolulu Sunrise
22 January 2024

David “Kawika” L. Morse
CISSP, MBA-Information Systems



TOC

“totally amazing concepts!”

[What is Artificial Intelligence?](#)

[A Brief History of AI...](#)

[Who uses AI today?](#)

[How can I use AI?](#)

[What are the challenges?](#)

[For students and researchers...](#)



Example AI generated image.

More Information at openmorse.org/ai

What is Artificial Intelligence?

AI is not a simulation of human intelligence, but may have similarities.

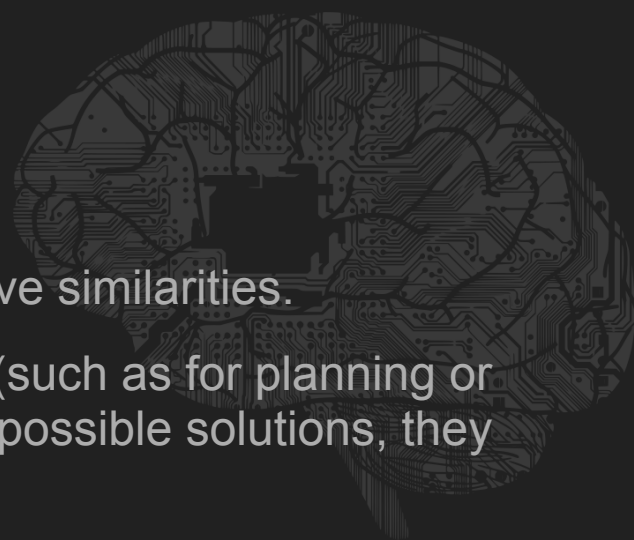
AI systems may use searches to solve many problems (such as for planning or game-playing algorithms). By searching through many possible solutions, they identify the most likely answers.

AI systems are trained via supervised, unsupervised or reinforcement learning.

Text-based generative AI systems are pre-trained on Large Language Models (LLMs) to predict the most likely next “tokens” and recognize patterns. When given an input prompt they construct new output via repeated prediction.

“On an operational level for business use, AI is a set of technologies that are based primarily on machine learning and deep learning, used for data analytics, predictions and forecasting, object categorization, natural language processing, recommendations, intelligent data retrieval, and more.”[1]

[See the AI meta article on Wikipedia here.](#)[2]



A Brief History of AI

Earliest “self learning” AI program written in the 1950s to play checkers.[3]

In the 1960s, advances in theory and programming improved AI learning, dialog and scientific reasoning, and AI robotics incorporate vision.

Expert systems (inference engine + knowledge base) used widely in 1980s from the space shuttle to mortgage loans. AI begins to leverage parallel computing.

The 1990s saw self-driving cars, autonomous consumer “pet” robots (AIBO), and Deep Blue beats the reigning world champion in chess.

By the early 2010s, Natural Language (NLP) advances allow Apple, Google and Microsoft to launch apps for answering questions and making recommendations.

Who uses AI today?

Everybody!

Consumer apps on phones and via the Web.

From customer service to medical diagnostics.

Robotics, self driving cars, and drones on the border.

Artists, authors, teachers, poets, playwrights and musicians.

Banking, investing, marketing, even farmers and environmentalists!

Businesses are using AI to “streamline processes, optimize costs, prevent human error, assist customers, manage IT systems, and alleviate repetitive tasks...”[4]



How can I use AI?

<u>Bard</u>	Google's generative AI chatbot.	Free "experimental". Critics say it is less capable than ChatGPT.
<u>ChatGPT</u>	OpenAI's generative AI chatbot.	Free and paid versions. Versatile: programming, music, business, translate, essays, plays, poetry, etc.
<u>CoPilot</u>	Microsoft's generative AI chatbot.	Versatile: poetry, cite sources, music, translate, separate Image Creator, integrations across products (e.g., Visual Studio for programming)
<u>Image Creator</u>	Microsoft's generative image chatbot	Uses DALL-E 3 from OpenAI. Limited quality, lacks accuracy (may not conform to prompts)
<u>Stable Diffusion</u>	Stability AI's generative image chatbot	A deep learning text to image model with open source code. Various tools and platforms and pricing.

Example generative text using ChatGPT (3.5)



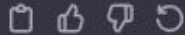
You

create a limerick about homemade pie and fish



ChatGPT

There once was a cook who'd apply
To bake a fine pie, oh my!
With fish and some spice,
It tasted so nice,
Homemade, it would make your heart fly!



You

what are the three most important concepts in cyber security?



ChatGPT

Three foundational concepts in cybersecurity are:

1. **Confidentiality:** This involves ensuring that sensitive information is accessible only to those who are authorized to access it. It encompasses practices like encryption, access controls, and data protection measures to prevent unauthorized access.
2. **Integrity:** This concept focuses on maintaining the accuracy and trustworthiness of data and systems. It involves measures to prevent unauthorized or unwanted changes, ensuring that data remains unchanged and reliable.
3. **Availability:** This refers to ensuring that systems and data are accessible and usable when needed by authorized users. Measures such as redundancy, backups, and disaster recovery plans are essential to maintain availability, even in the face of cyberattacks or system failures.



Example write a program using ChatGPT (3.5)

You

write a python program to convert temperature from Fahrenheit to Celsius



ChatGPT

Sure, here's a simple Python program that converts temperature from Fahrenheit to Celsius:

python Copy code

```
def fahrenheit_to_celsius(fahrenheit):
    celsius = (fahrenheit - 32) * 5/9
    return celsius

# Input temperature in Fahrenheit
fahrenheit = float(input("Enter temperature in Fahrenheit: "))

# Convert Fahrenheit to Celsius
celsius = fahrenheit_to_celsius(fahrenheit)

print(f"{fahrenheit} Fahrenheit is equal to {celsius:.2f} Celsius")
```

This program defines a function `fahrenheit_to_celsius` that takes a temperature in Fahrenheit as input and returns the equivalent temperature in Celsius using the conversion formula. Then it prompts the user to input a temperature in Fahrenheit, performs the conversion, and displays the result in Celsius.

You

update the code to include a unit test



ChatGPT

Certainly! Here's the updated code that includes capturing user input for the Fahrenheit temperature and then performs the conversion while also running the unit test:

python Copy code

```
import unittest

def fahrenheit_to_celsius(fahrenheit):
    celsius = (fahrenheit - 32) * 5/9
    return celsius

class TestTemperatureConversion(unittest.TestCase):
    def test_conversion(self):
        self.assertAlmostEqual(fahrenheit_to_celsius(32), 0)
        self.assertAlmostEqual(fahrenheit_to_celsius(212), 100)
        self.assertAlmostEqual(fahrenheit_to_celsius(98.6), 37)

if __name__ == "__main__":
    # Capture user input for Fahrenheit temperature
    fahrenheit = float(input("Enter temperature in Fahrenheit: "))

    # Convert Fahrenheit to Celsius
    celsius = fahrenheit_to_celsius(fahrenheit)
    print(f"{fahrenheit} Fahrenheit is equal to {celsius:.2f} Celsius")

    # Run unit test
    unittest.main()
```


Example Image Creation using MS Bing Image Creator

Prompt: (iterated) *“Hawaiian outrigger canoe on the ocean with octopus paddling. Only the octopus is in the outrigger canoe. And the octopus is using multiple paddles.”*

*Multiple iterations of prompt.

*Result does not conform to prompt (except perhaps for one image).



What are the challenges?

AI systems can be biased - if the data or the algorithms are biased, so is output. What if used in medical diagnosis, or in law enforcement or military operations?

AI is vulnerable to attack. Malicious actors can either poison the data (the learning material) or alter the inputs (the prompts) to achieve their goals. These attacks take advantage of algorithm weakness (versus programming issues).

AI systems can “hallucinate”, creating output that makes little sense. And large, complex AI systems can be “black box” - we don’t actually know how they work!

With large data sets come issues of privacy, accuracy, surveillance, and copyright.

Misinformation and reputational damage via “deepfake” images and video.

Creating AI “apps” requires little expertise, little time, little cost, and apps can be [integrated with third party apps and services](#) for transactions, data exchange, etc. 😲

Safety, Security, Privacy, Regulation

[Blueprint for an AI Bill of Rights](#) and [Executive Order on AI](#)

[US DHS, CISA's Roadmap for AI](#)

DOD [AI Training Resources](#) and the [Responsible AI Toolkit](#) (risk assessment and dev tools)

[EU Artificial Intelligence Act update](#)

[Copyright Registration Guidance: Works Containing Material Generated by AI](#)

[Applying lessons from traditional cybersecurity to emerging AI-model risks](#)

[Attacking Artificial Intelligence: AI's Security Vulnerability](#)

[10 best practices for securely developing with AI](#)

[OWASP AI Security and Privacy Guide](#) and [Top 10 Security issues for LLMs](#)

[AI hacking at DEF CON](#)

OWASP Top 10 for LLM Applications

LLM01

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02

Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03

Training Data Poisoning

LLM04

Model Denial of Service

LLM05

Supply Chain Vulnerabilities

LLM06

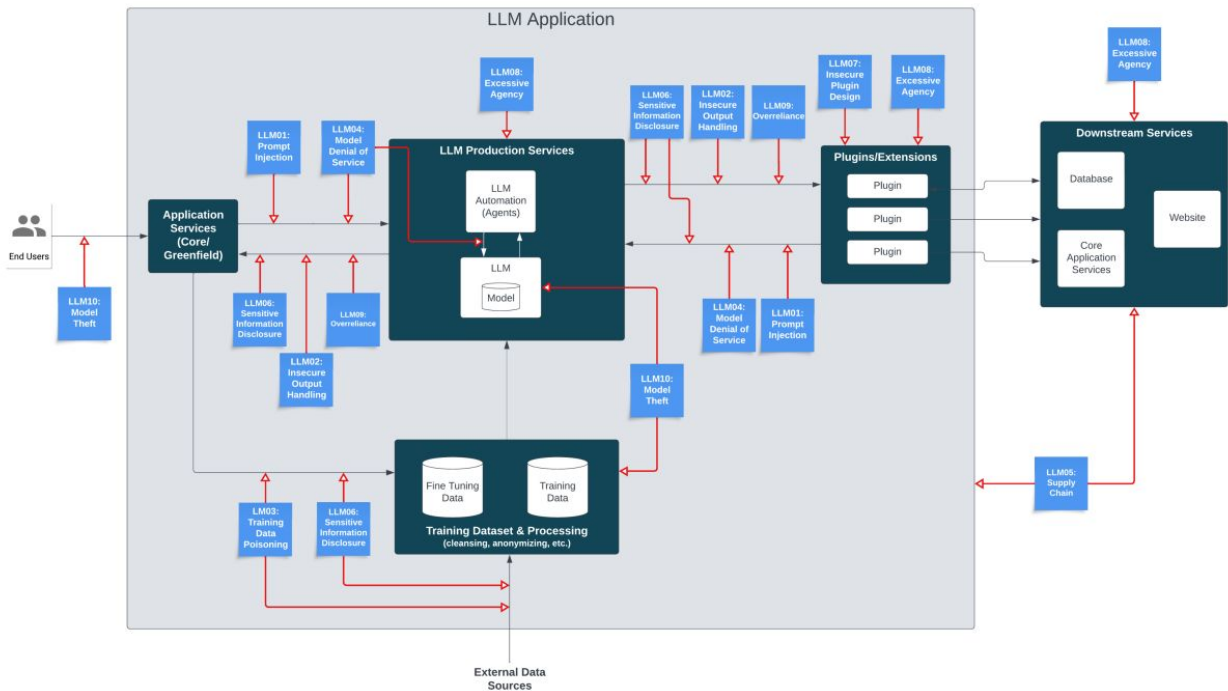
Sensitive Information Disclosure

LLMs may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. It's crucial to implement data sanitization and strict user policies to mitigate this.

LLM07

Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control. This lack of application control makes them easier to exploit and can result in consequences like remote code execution.



For Students, Developers, Researchers

There are multiple free (and paid) courses available online for AI/ML. Leverage online resources and work on getting [one or two solid certificates](#).

For those just beginning their technology journey, capture your work (even the simple stuff) in a GitHub portfolio. Learn the tools industry uses! And look for open source projects that you can contribute to - even just helping them write documentation, or adding a few simple test functions to their code.

Join local technology groups - check meetup and eventbrite.

Don't ignore "cyber security" - it is not a separate field, it is all intertwined. As you develop software (or configure systems, or engineer prompts) consider security.

Learning Resources

Google AI/ML: [Data Scientist / Machine Learning Engineer learning path](#). Note the [Professional Machine Learning Engineer](#) exam does not cover generative AI, as the tools used are evolving quickly. For generative AI, refer to the [Introduction to Generative AI Learning Path](#) (all audiences) or the [Generative AI for Developers Learning Path](#) (technical audience).

Microsoft (Azure): [AI community hub - courses and certificates, forums, events](#)

Amazon (AWS): [Machine Learning courses](#) (free and paid)

Harvard: [CS50's Introduction to Artificial Intelligence with Python](#) (free, or pay for cert)

freeCodeCamp: (example) [Azure AI Fundamentals cert prep](#), (example) [Machine learning with Python cert series](#)

Udemy: (example) [Artificial Intelligence A-Z 2023](#) (learn fundamentals and build with Python)

Coursera: Professional certificate tracks approximately \$40/month (e.g., IBM series)

References:

1. <https://cloud.google.com/learn/what-is-artificial-intelligence>
2. https://en.wikipedia.org/wiki/Artificial_intelligence
3. https://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence
4. <https://www.cio.com/article/652775/12-most-popular-ai-use-cases-in-the-enterprise-today.html>

More Information at openmorse.org/ai

Unless otherwise noted, all images are creative commons licensed and free to use.