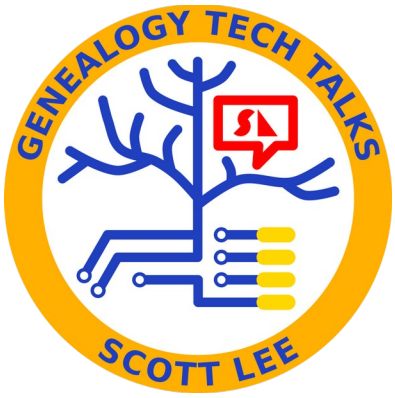
Artificial Intelligence And Genealogy

by Scott Lee

Scott@ScottLee.com

Presentation Handouts: https://GenealogyTechTalks.com/talks/ai

My website: www.GenealogyTechTalks.com



Learning To Use AI In Genealogy

How to Use ChatGPT for Genealogy the Right Way
May 2025, 8 min – https://youtu.be/Y1knsOtNYiU
Explains why ChatGPT is not a search engine, four main use-cases (summarizing, extracting, generating, translating), example prompts for genealogy, and how to handle hallucinations safely when working with family-history data.

Introduction to AI for Genealogy
Jan 2026, 1 hr 24 min – https://youtu.be/ujRSvRDCJVI
[Note, skip to 5:53 to avoid the count down timer] A 2026 library webinar by Peggy Jude that explains what AI is, why genealogists might use it, basic terminology, and practical examples like summarizing records and drafting research notes, with a strong emphasis on ethics and accuracy.

Beginning with AI Genealogy: Getting Started with Generative Artificial Intelligence in 2025
Apr 2025, 1 hr – https://youtu.be/rYWEl6OSIBE
FamilySearch's introductory class on using large language models (like ChatGPT and similar tools) for family history; covers what AI can and cannot do, what tools to start with, and how to set realistic expectations for using AI in research workflows.

Can You Use AI to Do Family History Research? - James Tanner (30 Mar 2025)
Apr 2025, 51 min – https://youtu.be/gIRLVvOmRBE
A conceptual talk about how AI fits into the traditional genealogy research cycle, what it can realistically automate (like full-text searching in records) and what still requires a human genealogist, useful if you want to understand "where AI fits" before using it.

Master AI Prompts for Breakthrough Genealogy Research
Aug 2025, 1 hr – https://youtu.be/gQe77lDyops
Shows how to use a simple 5-part prompt formula to get better results from AI tools, focusing on tasks like summarizing records, extracting key facts, and generating research ideas, aimed at genealogists who have basic AI familiarity.

Guidelines for the Responsible Use of AI in Genealogy
Dec 2025, 1 hr – https://youtu.be/Pupm38Uz2Io
Expert panel that outlines best practices for accuracy, privacy, disclosure, education, and compliance when integrating AI into family-history work.

A Deeper Dive Into AI Technology

Who Invented A.I.? - The Pioneers of Our Future
Dec 2019, 19 min – https://youtu.be/IBe2o-cZncU
A documentary-style overview of the history of AI from early neural-network attempts to modern deep learning, highlighting key figures like Frank Rosenblatt, Geoffrey Hinton, Yann LeCun, and the rise of systems like AlphaGo.

What The 2030s Will Look Like with Ray Kurzweil
Nov 2022, 11 min – https://youtu.be/ZJF6GoE-R8s
Astrophysicist Neil deGrasse Tyson interviews futurist Ray Kurzweil about the “singularity,” AI passing the Turing test, and how human intelligence might merge with machines in the 2030s.

What is Artificial Intelligence Exactly?
Jul 2016, 9 min – https://youtu.be/kWmX3pd1f10
Introduces core definitions of AI, different kinds of intelligence and expert systems, and explains terms like machine learning, neural networks, and strong vs. weak AI in clear, non-technical language.

But what is a neural network? | Deep learning chapter 1
Oct 2017, 19 min – https://youtu.be/aircAruvnKk
A visual, math-focused explanation of how basic neural networks work, using the example of recognizing handwritten digits to build intuition for layers, weights, biases, and “learning.”

Artificial Intelligence Glossary

This is a glossary of terms used in the presentation, *Artificial Intelligence And Genealogy*, plus other related terms. This was largely generated by AI, edited by Scott Lee.

3-D Face Reconstruction: An AI process mentioned as a near-future capability that uses multiple 2D photographs to build a three-dimensional model of a person's face.

Activation Function: In a neural network, an activation function is a mathematical equation that determines whether a neuron should be activated or not. It decides how much signal a neuron should pass forward.

AI (Artificial Intelligence): Computer systems or machines that perform tasks that typically require human intelligence, such as learning, decision-making, and language understanding.

Algorithm: A sequence of instructions that a computer follows to solve a problem or complete a task.

AlphaGo: An AI program developed by DeepMind that defeated the world champion in the board game Go.

AlphaZero: An AI program developed by DeepMind capable of mastering games like chess and Go.

ANI (Artificial Narrow Intelligence): AI focused on a single task or narrow set of tasks, unlike humans who can learn a wide range of tasks.

AGI (Artificial General Intelligence): A future form of AI that can understand, learn, and apply its intelligence broadly, similar to a human's cognitive abilities.

Automated Research Assistant: A future AI tool designed to mimic a professional genealogist by suggesting tailored research strategies and remembering where a user has already searched.

Backpropagation: An AI technique used to improve neural network performance by adjusting weights based on errors in output using differential calculus.

Bias (in AI): Inclination or prejudice in AI systems, often a result of the data on which they are trained.

Bias (in Neural Networks): In the context of a neural network, a 'bias' is a parameter associated with a neuron that adjusts the output along with the weighted sum of the inputs.

Chatbot: A computer program designed to simulate conversation with human users, especially over the Internet.

ChatGPT: An AI language model developed by OpenAI, trained on a large amount of text data to generate human-like text.

Claude: A Large Language Model (LLM) similar to ChatGPT that is trained on vast amounts of internet text to assist with research.

Data Cleansing: The process of correcting or removing inaccurate records from a dataset.

Deep Blue: An AI chess-playing computer developed by IBM that defeated world champion Garry Kasparov.

Deep Learning: A subset of machine learning involving neural networks with multiple layers that enable the learning of complex patterns.

DeepMind: An AI research laboratory (now owned by Google) responsible for creating advanced systems like AlphaGo and MuZero.

DeepSeek: A Large Language Model (LLM) similar to ChatGPT that is trained on vast amounts of internet text to assist with research.

Differential Multivariate Calculus: A branch of mathematics used in the Backpropagation process to calculate the "slope" of errors, allowing the system to adjust weighting factors more efficiently.

Discriminator: One half of a Generative Adversarial Network (GAN) responsible for detecting whether a generated image is a "fake" or "real".

ExaFLOPS: A measure of computer performance representing one quintillion (10^{18}, or a billion billion) floating-point operations per second; the predicted milestone for computing power in the coming decade.

Facial Recognition: AI technology that identifies individuals by analyzing facial features in images or videos.

FAN Club (Friends, Associates, Neighbors): A genealogical research strategy that examines an ancestor's wider social network.

Floating Point: A method used by computers to represent and process very large or very small numbers using a system similar to scientific notation (e.g., 1.23×10^9). Unlike "integer" math, which only uses whole numbers, floating point math is essential for the complex, high-precision calculations required for Neural Networks, and 3D graphics.

GAN (Generative Adversarial Network): AI systems where two neural networks compete with each other to improve their capabilities, often used in image generation.

Gemini: A Large Language Model (LLM) similar to ChatGPT that is trained on vast amounts of internet text to assist with research.

Generator: The component of a GAN that creates new data (such as pictures of faces) to attempt to "fool" the discriminator.

Genetic Analysis (in Genealogy): Using DNA testing to understand ancestry and familial relationships.

Go: An ancient board game originating from China, renowned for its deep strategy and complexity.

GPUs (Graphics Processing Units): Hardware devices optimized for handling complex calculations, essential in AI computations.

Hallucination: A phenomenon where a Large Language Model (LLM) generates information that sounds confident and plausible but is factually incorrect.

Handwriting Recognition: AI's ability to interpret and convert handwritten text into digital data.

Huang, Jensen: The CEO of NVIDIA who predicted a million-fold increase in computing power over the next ten years.

Kasparov, Garry: A world-class chess champion defeated by the AI system Deep Blue.

Ke Jie: A world-class professional Go player from China defeated by the AI system AlphaGo.

Large Language Models (LLMs): AI models that process, understand, and generate human language, trained on vast text datasets.

Machine Learning (ML): A subset of AI where machines learn from data and improve their performance over time.

MuZero: An advanced AI program by DeepMind, capable of learning and mastering multiple games without prior knowledge of their rules.

Neural Network: A computational model used in AI, structured similarly to the human brain, to process complex data patterns.

Next Word Prediction: The core mechanism of LLMs, where the model determines the “most probable word” to follow a sequence of text rather than understanding “truth”.

NVIDIA: A technology company that produces GPUs, the hardware originally meant for video games that now provides the massive power required to train AI.

Photo Dating: Using AI to estimate the date of a photograph by analyzing clothing styles, background objects, and the estimated age of the subjects.

Predictive Algorithms: AI algorithms used to predict future outcomes based on historical data.

Record Matching: An AI process in genealogy where historical records are matched to individuals in family trees.

Standardization (in Data): Making data uniform in format and structure, often used in genealogical records to maintain consistency.

TeraFLOP (Trillion Floating Point Operations Per Second): A measure of computer performance, particularly relevant in AI for processing large datasets quickly.

Tic-Tac-Toe: A simple game used as an example in explaining AI and neural networks.

Training (in AI): The process of teaching an AI system by feeding it data and allowing it to adjust and improve its performance.

Turing Test: Named after the British mathematician and computer scientist Alan Turing, it is a measure of a machine’s ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human.

Weighting Factors (Weights): Numerical values assigned to inputs within a Neural Network that determine how much influence a specific signal has on a neuron's output.

YottaFLOP: A measure of computer performance representing 10^{24} operations per second. This is currently the highest named unit of computing speed in the International System of Units (SI) and is 1,000 times faster than a ZettaFLOP.

ZettaFLOP: A measure of computer performance representing 10^{21} operations per second. This is 1,000 times faster than an ExaFLOP and 1,000,000 times faster than a PetaFLOP.

Computing Power: Scale of Magnitude

TeraFLOP	10^{12} (Trillion)	1996 Supercomputer / Modern GPU
PetaFLOP	10^{15} (Quadrillion)	1,000 times faster than a TeraFLOP
ExaFLOP	10^{18} (Quintillion)	Current state-of-the-art supercomputers (2022+)
ZettaFLOP	10^{21} (Sextillion)	1,000 times faster than a ExaFLOP
YottaFLOP	10^{24} (Septillion)	1,000 times faster than an ZettaFLOP