

# Large Language Models for Software Engineering

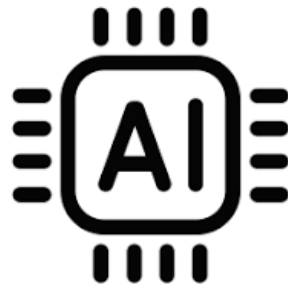
Eduardo Figueiredo

<http://www.dcc.ufmg.br/~figueiredo>

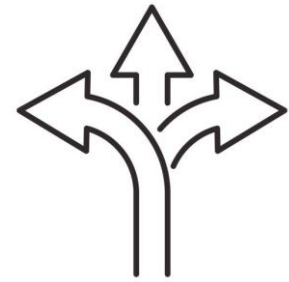
# [ A Typical AI System ]



Data



Model



Decision

- Data are used to train the model
  - Decisions are made for new data based on the patterns identified in the training data

# Structure vs. Unstructured Data



Data

## Structured Data

id	name	email
20244123	Thomas Anderson	neo@gmail.com
20251234	Sarah Connor	sarah@gmail.com
20264321	Marty McFly	mcfly@gmail.com

## Unstructured Data



Arthur B. ✓  
@ArthurB

Follow



You're going to ditch Salesforce for an alternative distributed by Microsoft or Google who previously couldn't justify the development cost.



Justine Moore ✓ @venturetwins · Feb 6

I love how everyone is saying "SaaS is dead" like you're going to get the Fortune 500 to ditch Salesforce for a CRM vibecoded by a 13-year-old

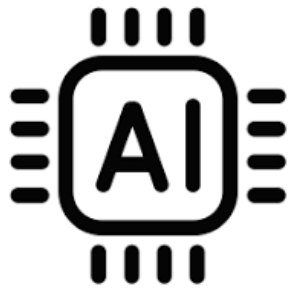
# Structure vs. Unstructured Data



Data

- Structured Data
  - Often quantitative (numbers)
  - 👍 Easy to analyze
  - 👎 Limited insights
- Unstructured Data
  - Often qualitative (text)
  - 👎 Difficult to analyze
  - 👍 Meaningful insights

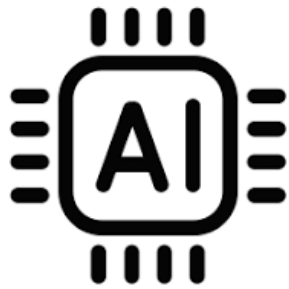
# Categories of AI Models



Model

- Supervised models
  - Trained on a set of labeled data
  - Make decisions for unseen data
- Unsupervised models
  - There is no training
  - Analyze the data to find patterns
  - Group similar data points

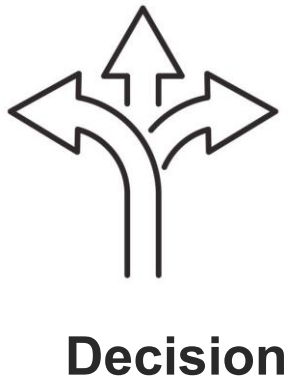
# Examples of AI Models















Model

- Supervised models
  - *Linear Regression*: predicts the next value
  - *Decision Tree*: predicts the best category (classifier)
- Unsupervised models
  - *K-means*: groups unlabeled data (clustering)

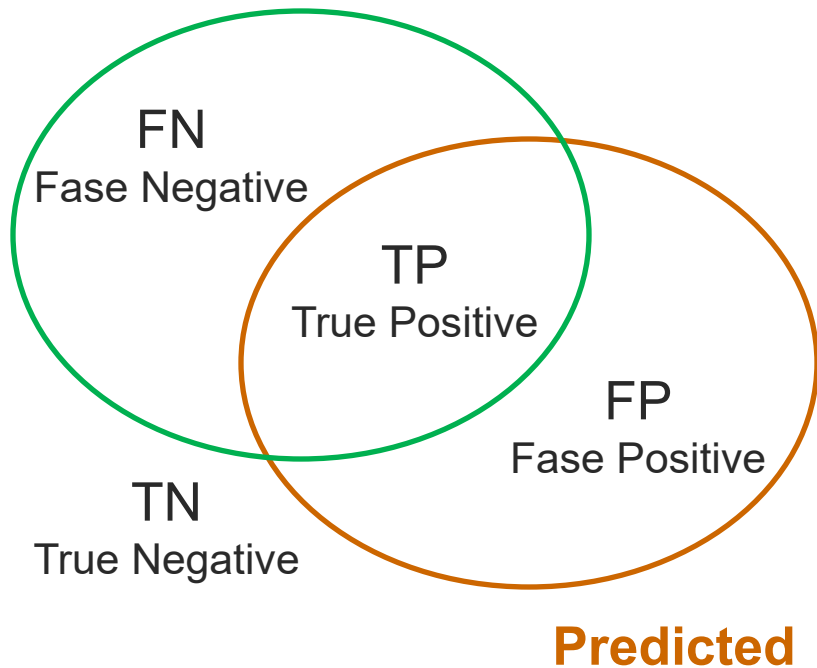
# [ Making a Decision ]



	Actual	Predicted	
			TP True Positive
			FN False Negative
			FP False Positive
			TN True Negative

# Measuring Effectiveness

Actual



## ■ Precision

- How small is FP?
- $TP / (TP+FP)$

## ■ Recall

- How small is FN?
- $TP / (TP+FN)$

# [ F1 Score and Accuracy ]

- F1 Score (F-Measure)


- The harmonic mean of precision and recall.

$$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Accuracy

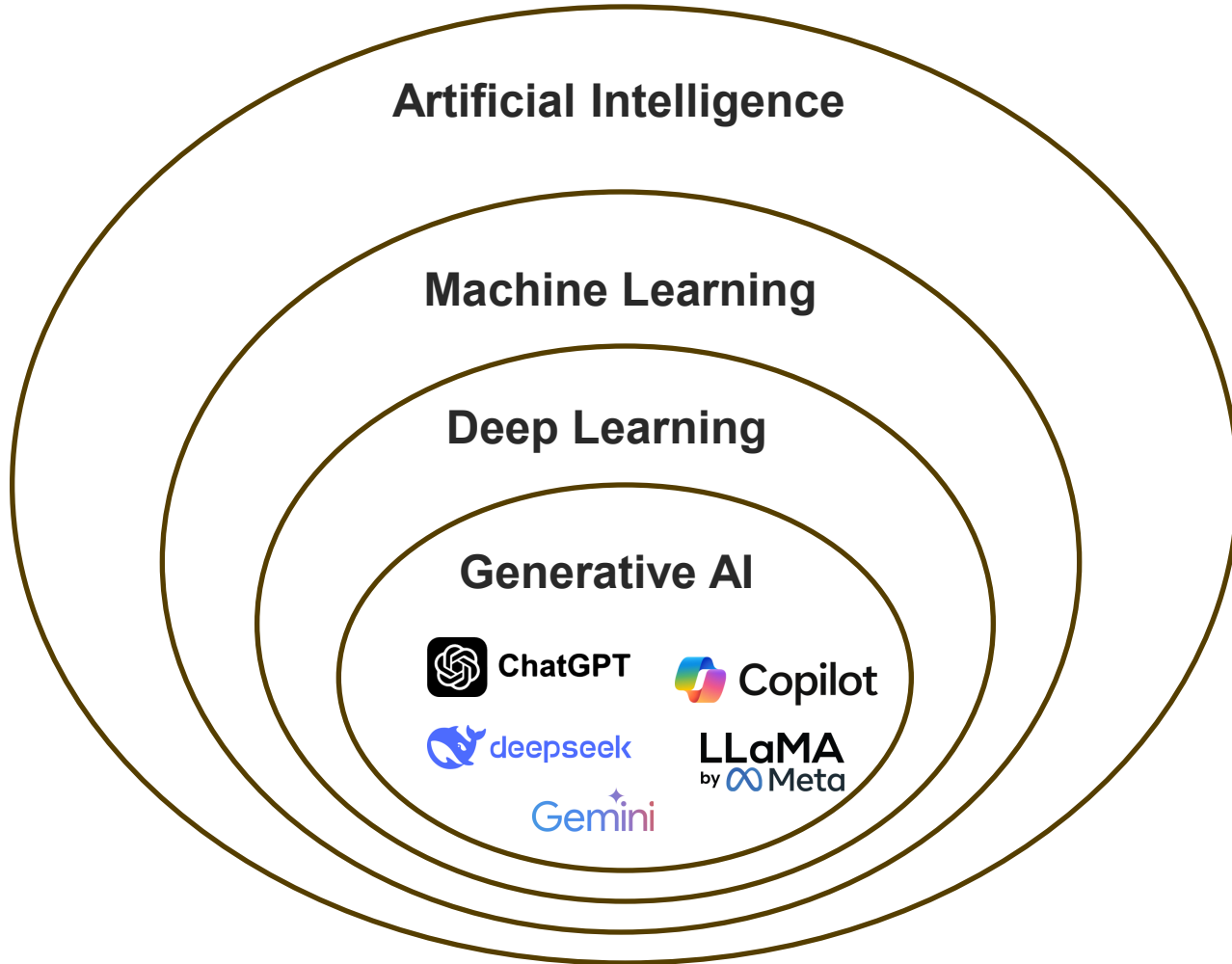
- It is the overall accuracy, but can be misleading in imbalanced datasets.

$$\frac{TP + TN}{TP + TN + FP + FN}$$



# Large Language Models

# Generative AI and LLMs

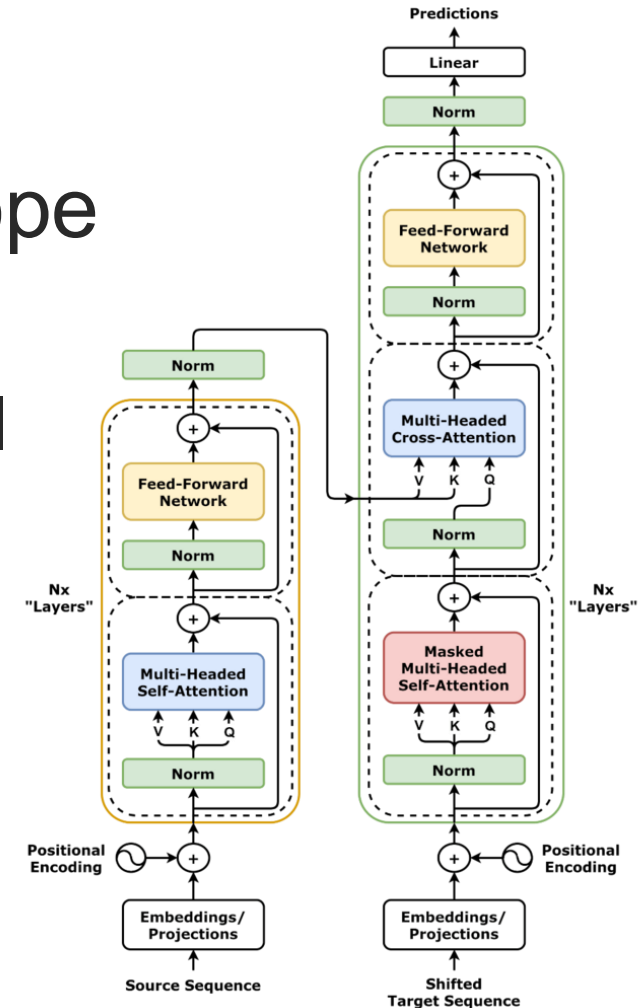


# [ Large Language Models (LLMs) ]

- Every LLM is a form of Generative AI
  - Generative AI can create new content rather than just analyzing existing data
  - An LLM learns patterns from massive datasets to generate original outputs
- Transformer allows LLMs to “pay attention” to the relationships between words

# Transformer Architecture

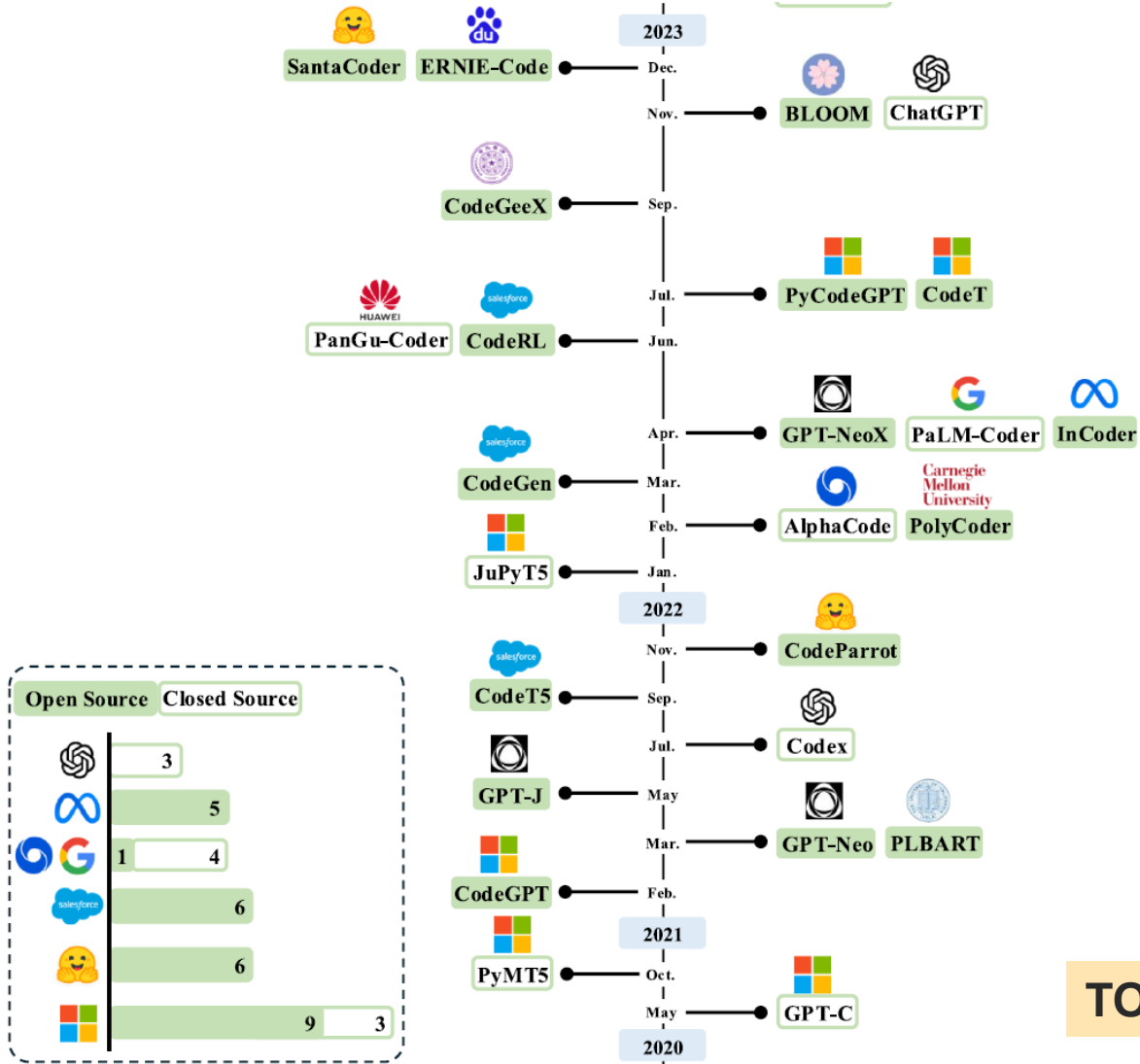
- At each layer, each token is contextualized within the scope of the context window
  - Text is converted to numerical representations (tokens)
  - Each token is converted into a vector
- Later variations are adopted for training LLMs



# [ AI for Software Engineering ]

- The main expected benefit is automation
  - By applying AI across the entire lifecycle in partnership with human developers
- AI Tools in Software Development
  - *Claude Code*: terminal-based tool built for agent-style, multi-step workflows
  - *GitHub Copilot*: assistant for real-time code suggestions and chat-based help

# LLMs for Code Generation (1)





# Examples of LLMs



# [ OpenAI GPT and ChatGPT ]

- GPT: Generative Pre-trained Transformer
  - ChatGPT (front-end) is an application built on GPT models



Model	Year	Size	Highlight
GPT-1	2018	117M	Proof of concept
GPT-2	2019	1.5B	Text generation
GPT-3	2020	175B	Widely adopted
GPT-4	2023	n/a	Improved reasoning
GPT-5	2025	n/a	Agentic AI

# [ GitHub Copilot ]

- An AI-powered coding assistant integrated into IDEs like VS Code
  - It was developed by GitHub and OpenAI

- Key Features

- Code completion
- Function generation
- Inline suggestions
- Multi-language support



# [ Google Gemini ]

- Similar to ChatGPT, it is a multimodal LLM system from Google
  - Gemini can process text, images, code, audio, etc.



- Key Features
  - Text, image, video understanding
  - Strong reasoning capabilities
  - Integration with Google ecosystem

# [ LLaMA ]

- An open LLM family from Meta
  - Small Models (7B), Medium Models (70B)
  - Strong performance with fewer parameters
- Key Features
  - Lightweight compared to GPT
  - Can run locally (important for privacy)
  - Strong research adoption



# [ DeepSeek ]

- A Chinese high-performance LLM family focused on reasoning and coding
  - It relies on Mixture of Experts (MoE)
  - Not all parameters are used at once



- Key Features
  - Strong math and coding abilities
  - Efficient training techniques
  - Competitive with top-tier models

# [ Claude ]

- Claude is a family of LLMs by Anthropic
  - Its purpose is similar to ChatGPT, Gemini...
  - An emphasis on AI alignment and safety

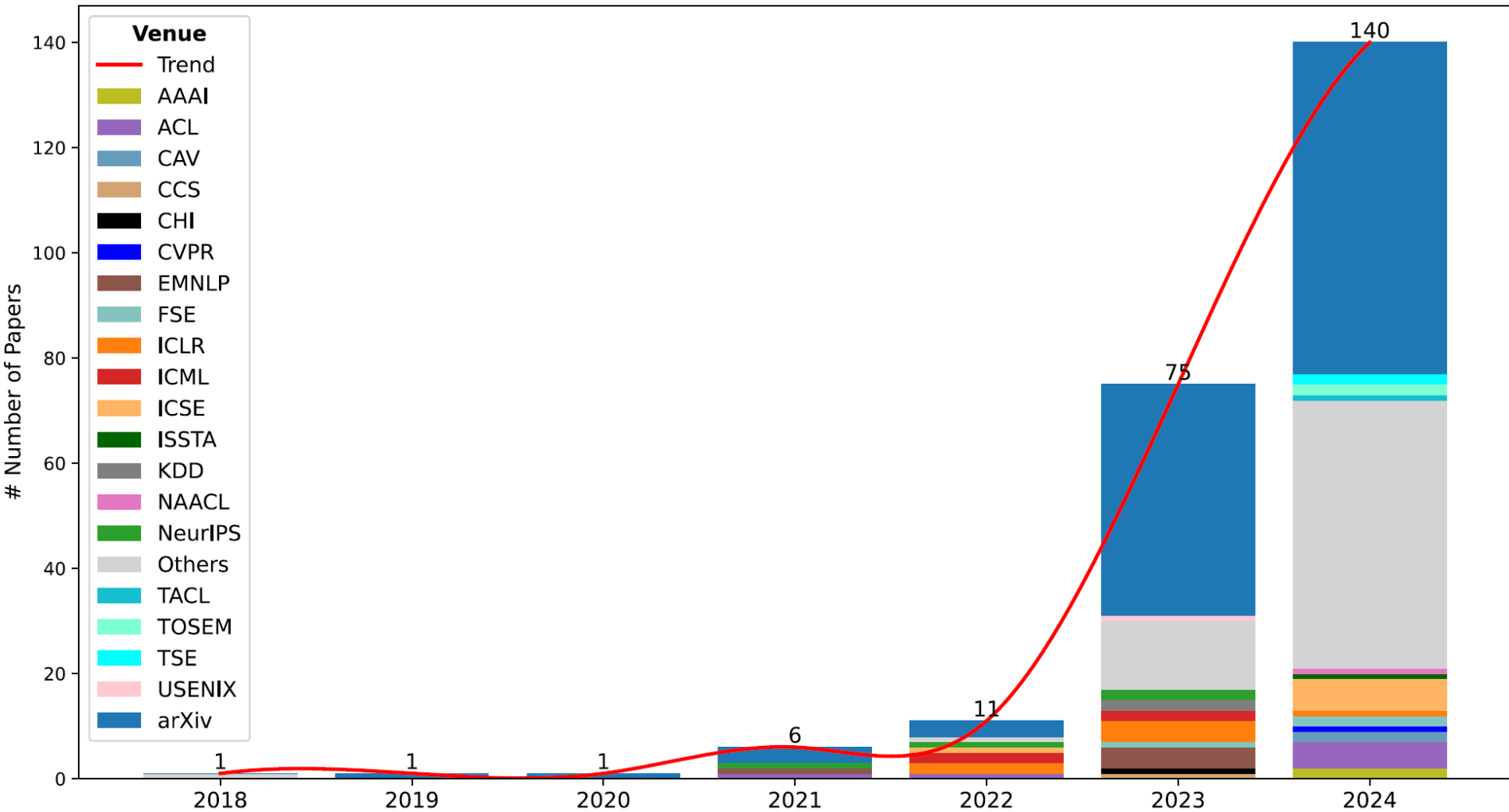
- It has three main variants
  - *Haiku*: speed and efficiency
  - *Sonnet*: balance
  - *Opus*: maximum capability

 **Claude**

# Comparison of LLMs

<b>Model</b>	<b>Release Year</b>	<b>Current Version</b>	<b>Approx. Size</b>
GPT	2018	GPT-5	n/a
Copilot	2021	GPT-5	n/a
Gemini	2023	Gemini 3.1	n/a
LLaMA	2023	LLaMA 4	7 to 70B
DeepSeek	2023	DeepSeek-R1	7 to 236B
Claude	2023	Claude 3	n/a

# [ LLM for Code Generation Research ]



# Will Software Engineers survive?

FINANCIAL TIMES

Opinion **The AI Shift**

## The AI Shift: Will software engineers survive agentic AI?

A data deep dive shows that job vacancies are rising — but only for senior developers

SARAH O'CONNOR



© Efi Chalikopoulou

① We're showing you this article with reduced functionality due to intermittent issues on FT.com. This article is up-to-date as of March 26 2026.

John Burn-Murdoch and Sarah O'Connor MARCH 26 2026 12:30 PM

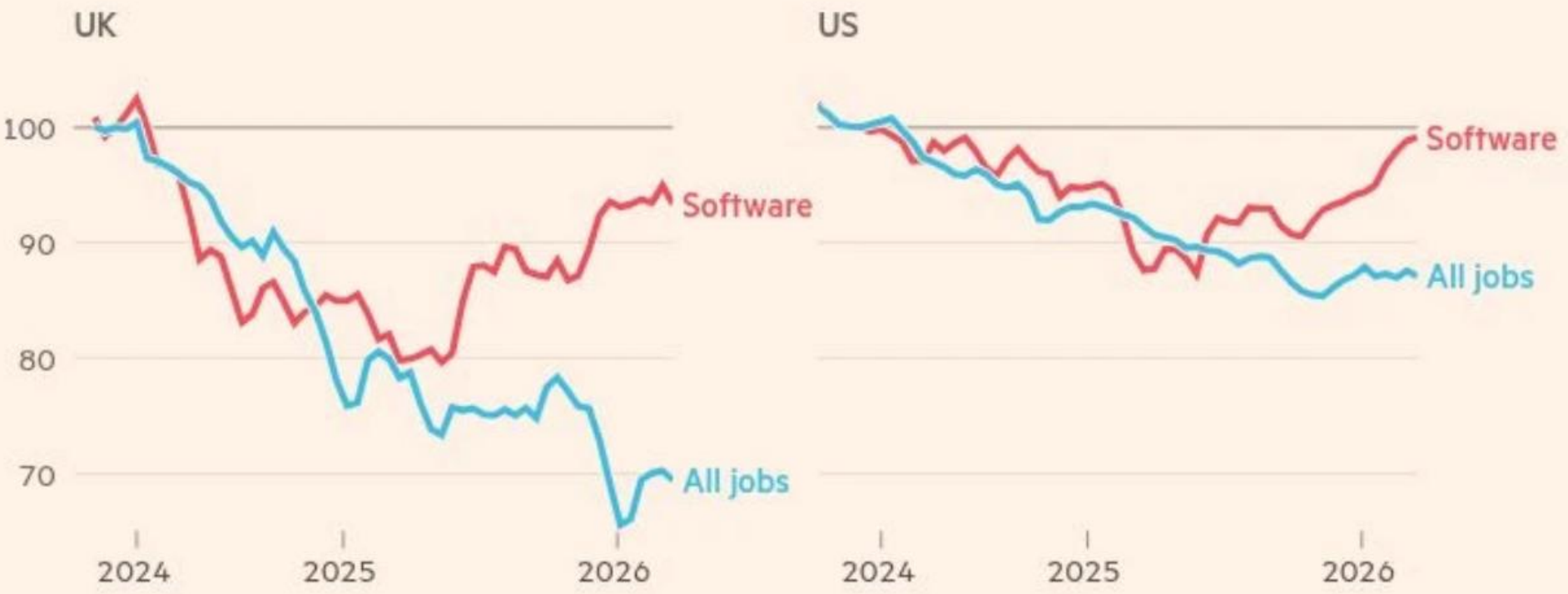
- Analyzed “millions of job adverts”
  - Job posts for SE is increasing
  - Job posts for junior SE is decreasing

*“As software becomes quicker and cheaper to build, people just want more of it.”*

Source: [Financial Times](#)

# [ Job posts for SE are increasing... ]

Job postings volume (Mar 2024 = 100)

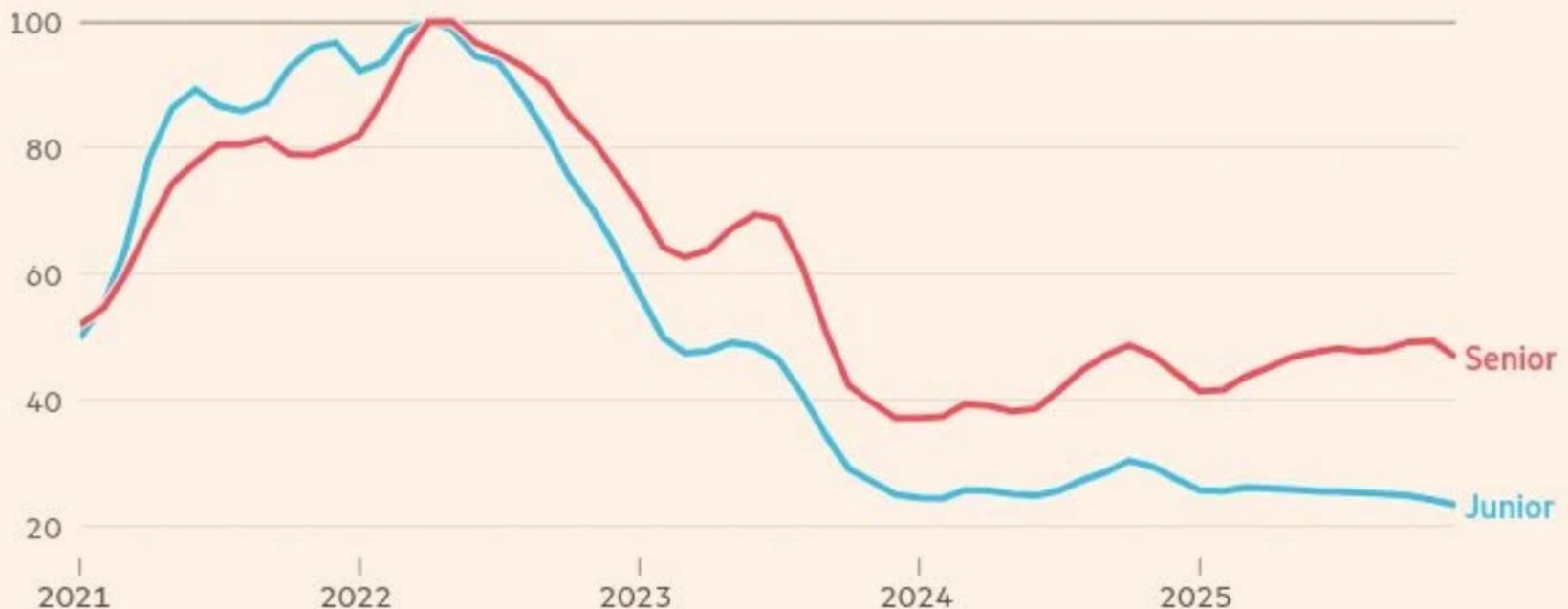


Source: FT analysis of [Indeed Hiring Lab data](#)  
FT graphic: John Burn-Murdoch / @jburnmurdoch  
©FT

# [ ...Increasing for Senior Developers ]

The uptick in software vacancies is exclusive to senior roles. Openings for junior developers remain scarce

Relative change in software job postings, by seniority of role (Apr 2022 = 100)



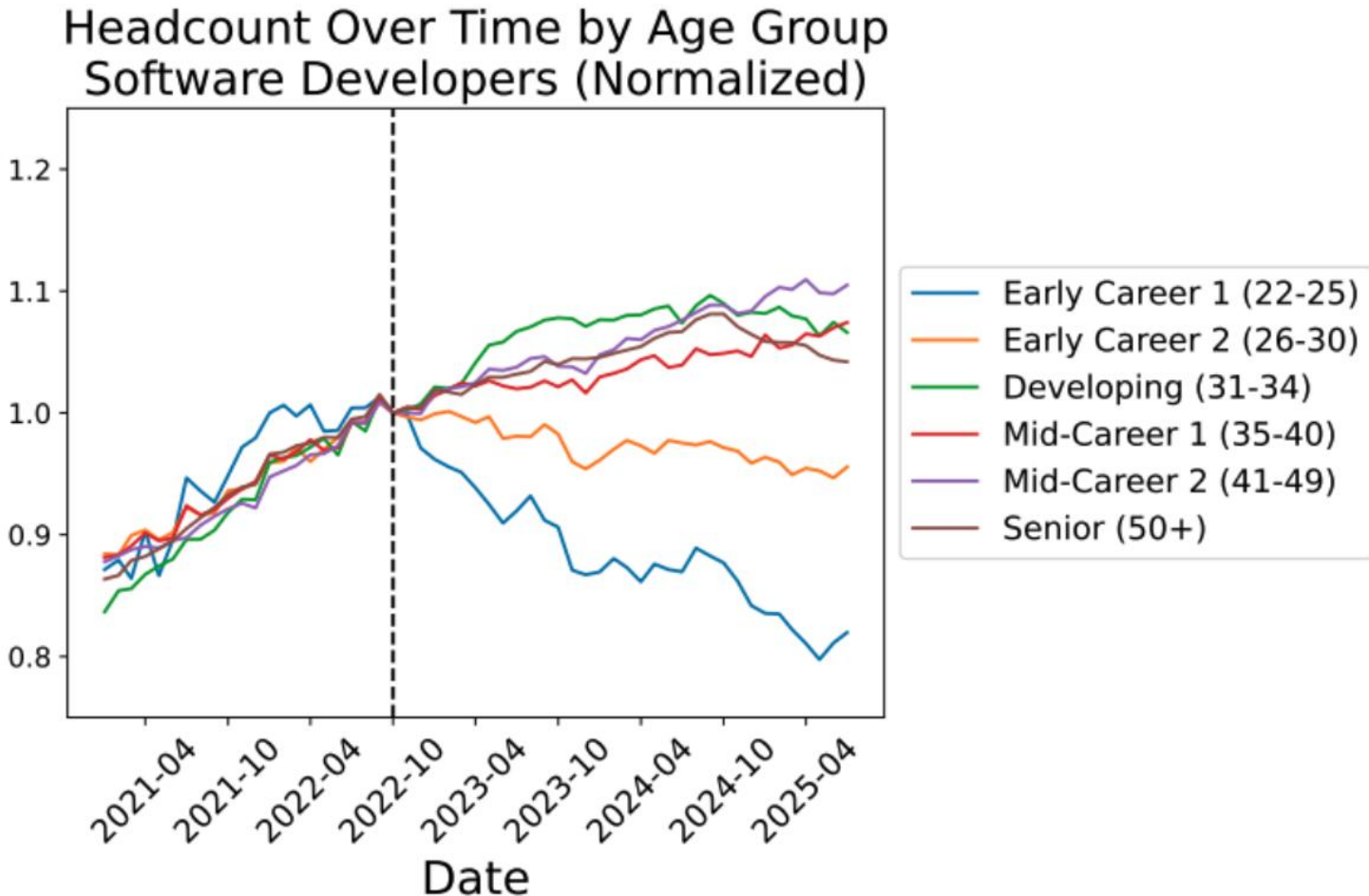
Source: FT analysis of [Lightcast job postings data](#). Junior = 0-2 years of experience, senior = 6+ years

FT graphic: John Burn-Murdoch / @jburnmurdoch

©FT

Source: [Financial Times](#)

# Similar Results by Stanford



# [ Bibliography ]

- Juyong Jiang *et al.* **A Survey on Large Language Models for Code Generation.** TOSEM, 2026.
- Ashish Vaswani *et al.* **Attention is all you need.** Advances in Neural Information Processing Systems (NIPS) 30, 2017.