

NOKIA

# Three (more) common networking misconceptions

Andrew Bender



Get to Fast Faster 2025

# AI in two dimensions

## AI for networks

Install, configure, operate, automate, optimize, secure, monitor, and troubleshoot

Today

In the Future



Fiber Networks



Wireless Networks – Wi-Fi



Wireless Networks – 5G



## Networks for AI

Build networks for the new era where computers work for all of us

...the way we work with each other

Fiber Networks



Wireless Networks – Wi-Fi



Wireless Networks – 5G





1

“AI is a cloud application...”

phi4

```
ollama run phi4
```

4.6M Downloads Updated 8 months ago

Phi-4 is a 14B parameter, state-of-the-art open model from Microsoft.

14b

Models [View all](#)

Name	Size	Context	Input
phi4:latest	9.1GB	16K	Text
phi4:14b <span>latest</span>	9.1GB	16K	Text

Readme

# 9.1GB

Phi-4 is a 14B parameter, state-of-the-art open model built upon a blend of synthetic datasets, data from filtered public domain websites, and acquired academic books and Q&A datasets.

gemma3

ollama.com/library/gemma3

Models GitHub Discord Turbo Search models Sign in Download

## gemma3

ollama run gemma3

16.2M Downloads Updated 1 month ago

The current, most capable model that runs on a single GPU.


vision 270m 1b 4b 12b 27b

### Models [View all](#)

Name	Size	Context	Input
gemma3:latest	3.3GB	128K	Text, Image
gemma3:270m	292MB	32K	Text
gemma3:1b	1.1GB	32K	Text
gemma3:4b <span>latest</span>	3.3GB	128K	Text, Image
gemma3:12b	8.1GB	128K	Text, Image
gemma3:27b	12.8GB	128K	Text, Image

**17GB**

Readme



The screenshot shows the Ollama website interface for the **gpt-oss** model. At the top, there's a navigation bar with links for Models, GitHub, Discord, and Turbo, along with a search bar and buttons for Sign in and Download. The main content area features the model name **gpt-oss** and a terminal snippet `ollama run gpt-oss`. Below this, it indicates 2.3M Downloads and an update from 1 month ago. A description states it's an OpenAI-style model for reasoning and agentic tasks. There are tags for 'tools', 'thinking', '20b', and '120b'. A 'Models' section contains a table with columns for Name, Size, Context, and Input. The table lists three variants: **gpt-oss:latest** (14GB), **gpt-oss:20b** (20GB), and **gpt-oss:120b** (65GB). A large black circle highlights the '65GB' size in the table, and a larger '65GB' text is overlaid below it. At the bottom, there's a blue banner with the 'gpt-oss' text and the OpenAI logo.

Name	Size	Context	Input
<b>gpt-oss:latest</b>	14GB	128K	Text
<b>gpt-oss:20b</b> <small>latest</small>	20GB	128K	Text
<b>gpt-oss:120b</b>	65GB	128K	Text

ollama.com/library/qwen3

Models GitHub Discord Turbo Search models Sign in Download

# qwen3

ollama run qwen3

8.1M Downloads Updated 1 month ago

Qwen3 is the latest generation of large language models in Qwen series, offering a comprehensive suite of dense and mixture-of-experts (MoE) models.

tools thinking 0.6b 1.7b 4b 8b 14b 30b 32b 235b

## Models

[View all](#)

Name	Size	Context	Input
qwen3:latest	5.2GB	40K	Text
qwen3:0.6b	523MB	40K	Text
qwen3:1.7b	1.4GB	40K	Text
qwen3:4b	2.5GB	256K	Text
qwen3:8b <span>latest</span>	5.2GB	40K	Text
qwen3:14b	9.3GB	40K	Text
qwen3:30b	19GB	256K	Text
qwen3:32b	20GB	40K	Text
qwen3:235b	142GB	256K	Text

142GB

deepseek-v3.1

ollama.com/library/deepseek-v3.1

Models GitHub Discord Turbo Search models Sign in Download

## deepseek-v3.1

ollama run deepseek-v3.1

47.3K Downloads Updated 2 weeks ago

DeepSeek-V3.1 is a hybrid model that supports both thinking mode and non-thinking mode.

tools thinking 671b


### Models

View all

Name	Size	Context	Input
deepseek-v3.1:latest	404GB	160K	Text
deepseek-v3.1:671b <span>latest</span>	404GB	160K	Text

## 404GB

### Readme

 deepseek

This model requires Ollama v0.11.7 which is currently in pre-release.

DeepSeek-V3.1 is a hybrid model that supports both thinking mode and non-thinking mode. Compared to the previous version, this upgrade brings improvements in multiple aspects:

Hybrid thinking mode: One model supports both thinking mode and non-thinking mode by changing the

2

“AI will be the driver for downstream traffic demand...”

NokiaGPT

gpt.nokia.com

**NOKIA GPT**

- New Chat
- Files
- Prompts
- Agents
- Chat History

### Explore with AI and Web Insights

Provides information from LLMs and includes relevant online results when web search is enabled.

**Content Creation**  
Develop engaging and effective (specific content format, e.g., emails or presentations) for (target audience) on the topic of (topic)

**Document Analysis**  
Provide a comprehensive analysis and highlight the key points from the attached document. (Attach or select document)

**Summarizing Content**  
Summarize the provided content, highlighting its key ideas and generating insightful questions for further exploration. (Paste text here)

**Coding Assistance**  
Generate a code snippet in (language) to accomplish the following task: (task)

**Research**  
Conduct a comprehensive analysis of the key concepts and definitions related to (topic)

**Translation**  
Translate the following paragraph from (source language) to (target language). (Please share the text)

P | I

+ Secure Web Search Base Work

100/100 messages left for today, GPT20b training data stops in 6/2024. Please refer to [T&C](#).  
NokiaGPT shall not be used to upload, process, or store the [U.S. Government data and/or U.S. Customer sensitive data and/or NSA Protected Information](#).

The screenshot shows the NokiaGPT web application interface. The browser address bar displays 'gpt.nokia.com'. The left sidebar contains navigation options: 'New Chat', 'Files', 'Prompts', 'Agents', 'Chat History', and 'User Guide'. The main content area is titled 'Explore with AI and Web Insights' and includes a sub-note: 'Provides information from LLMs and includes relevant online results when web search is enabled.' Below this, six AI capabilities are presented in a grid:

- Content Creation:** Develop engaging and effective (specific content format, e.g., emails or presentations) for (target audience) on the topic of (topic)
- Document Analysis:** Provide a comprehensive analysis and highlight the key points from the attached document. (Attach or select document)
- Summarizing Content:** Summarize the provided content, highlighting its key ideas and generating insightful questions for further exploration. (Paste text here)
- Coding Assistance:** Generate a code snippet in (language) to accomplish the following task: (task)
- Research:** Conduct a comprehensive analysis of the key concepts and definitions related to (topic)
- Translation:** Translate the following paragraph from (source language) to (target language). (Please share the text)

At the bottom, there is a text input field with a cursor, a '+', a 'Secure' dropdown, 'Web Search', 'Base Work' buttons, and an upward arrow. A status bar at the very bottom indicates '100/100 messages left for today. GPT20b training data stops in 6/2024. Please refer to [T&C](#). NokiaGPT shall not be used to upload, process, or store the [U.S. Government data and/or U.S. Customer sensitive data and/or NSA Protected Information](#).'

# Math time

# Math time

## Prompt

Please write a 3 page paper about the AI-related contributions of Bell Labs researcher Claude Shannon

# Math time

Prompt = 86 characters (22 tokens)

Please write a 3 page paper about the AI-related contributions of Bell Labs researcher Claude Shannon

# Math time

Prompt = 86 characters (22 tokens)

Please write a 3 page paper about the AI-related contributions of Bell Labs researcher Claude Shannon

## Response

### **Claude Shannon and the Foundations of Artificial Intelligence**

#### Abstract

Claude Elwood Shannon, often hailed as the “father of information theory,” made seminal contributions that reverberate through modern artificial intelligence (AI). While his most celebrated work—*A Mathematical Theory of Communication* (1948)—established the quantitative framework for data transmission, his insights into entropy, coding, and Boolean algebra laid the groundwork for machine learning, neural networks, and computational intelligence. This paper traces Shannon’s

# Math time

Prompt = 86 characters (22 tokens)

Please write a 3 page paper about the AI-related contributions of Bell Labs researcher Claude Shannon

Response = 5546 Characters (1387 tokens)

## **Claude Shannon and the Foundations of Artificial Intelligence**

### Abstract

Claude Elwood Shannon, often hailed as the “father of information theory,” made seminal contributions that reverberate through modern artificial intelligence (AI). While his most celebrated work—*A Mathematical Theory of Communication* (1948)—established the quantitative framework for data transmission, his insights into entropy, coding, and Boolean algebra laid the groundwork for machine learning, neural networks, and computational intelligence. This paper traces Shannon’s

# Math time

# Math time

Prompt Tokens = 22

## Math time

Prompt Tokens = 22

Response Tokens = 1387

## Math time

Prompt Tokens = 22

Response Tokens = 1387

Prompt Tokens

---

Prompt Tokens + Response Tokens

## Math time

Prompt Tokens = 22

Response Tokens = 1387

$$\frac{\text{Prompt Tokens}}{\text{Prompt Tokens} + \text{Response Tokens}} = \frac{22}{22 + 1387}$$

## Math time

Prompt Tokens = 22

Response Tokens = 1387

$$\frac{\text{Prompt Tokens}}{\text{Prompt Tokens} + \text{Response Tokens}} = \frac{22}{22 + 1387} \quad \text{Upstream Fraction} = 1.6\%$$

## Math time

Prompt Tokens = 22

Response Tokens = 1387

$$\frac{\text{Prompt Tokens}}{\text{Prompt Tokens} + \text{Response Tokens}} = \frac{22}{22 + 1387} \quad \text{Upstream Fraction} = 1.6\%$$

$$\frac{\text{Response Tokens}}{\text{Prompt Tokens} + \text{Response Tokens}}$$

## Math time

Prompt Tokens = 22

Response Tokens = 1387

$$\frac{\text{Prompt Tokens}}{\text{Prompt Tokens} + \text{Response Tokens}} = \frac{22}{22 + 1387} \quad \text{Upstream Fraction} = 1.6\%$$

$$\frac{\text{Response Tokens}}{\text{Prompt Tokens} + \text{Response Tokens}} = \frac{1387}{22 + 1387}$$

## Math time

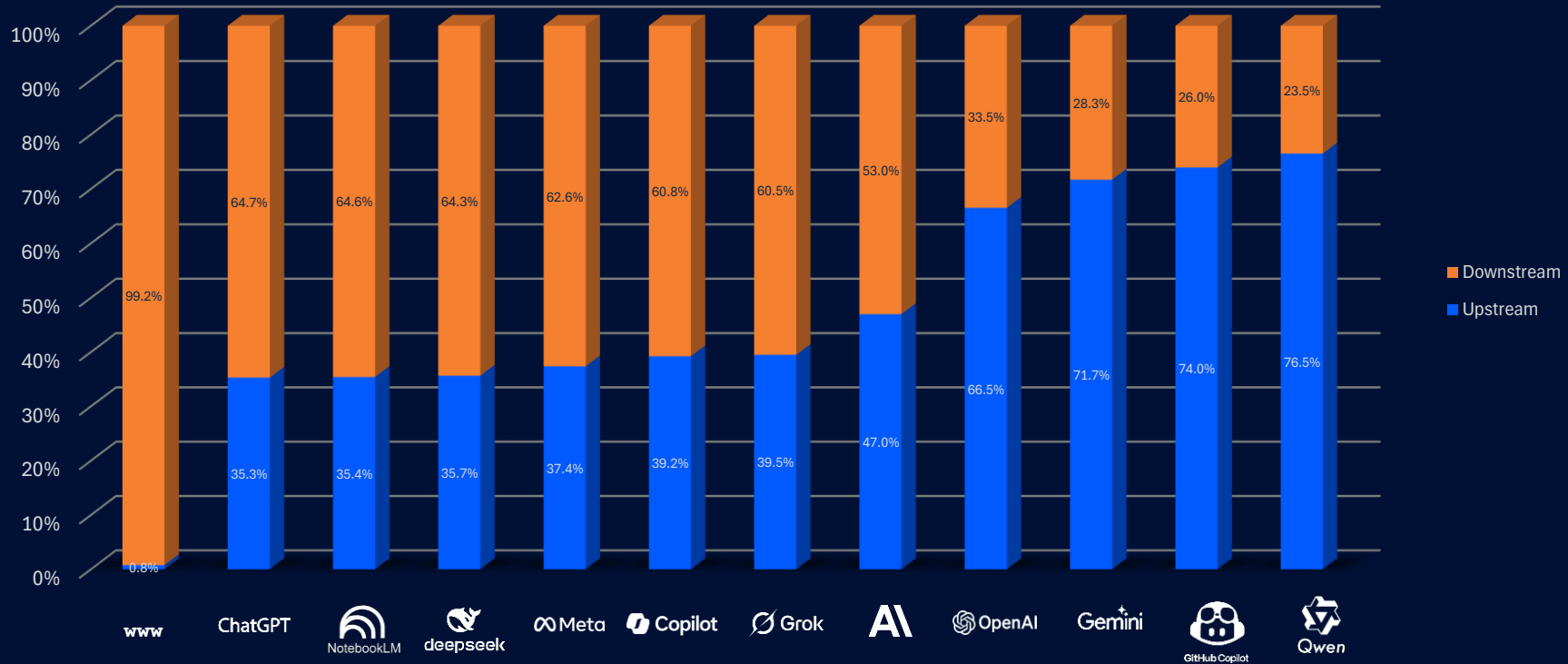
Prompt Tokens = 22

Response Tokens = 1387

$$\frac{\text{Prompt Tokens}}{\text{Prompt Tokens} + \text{Response Tokens}} = \frac{22}{22 + 1387} \quad \text{Upstream Fraction} = 1.6\%$$

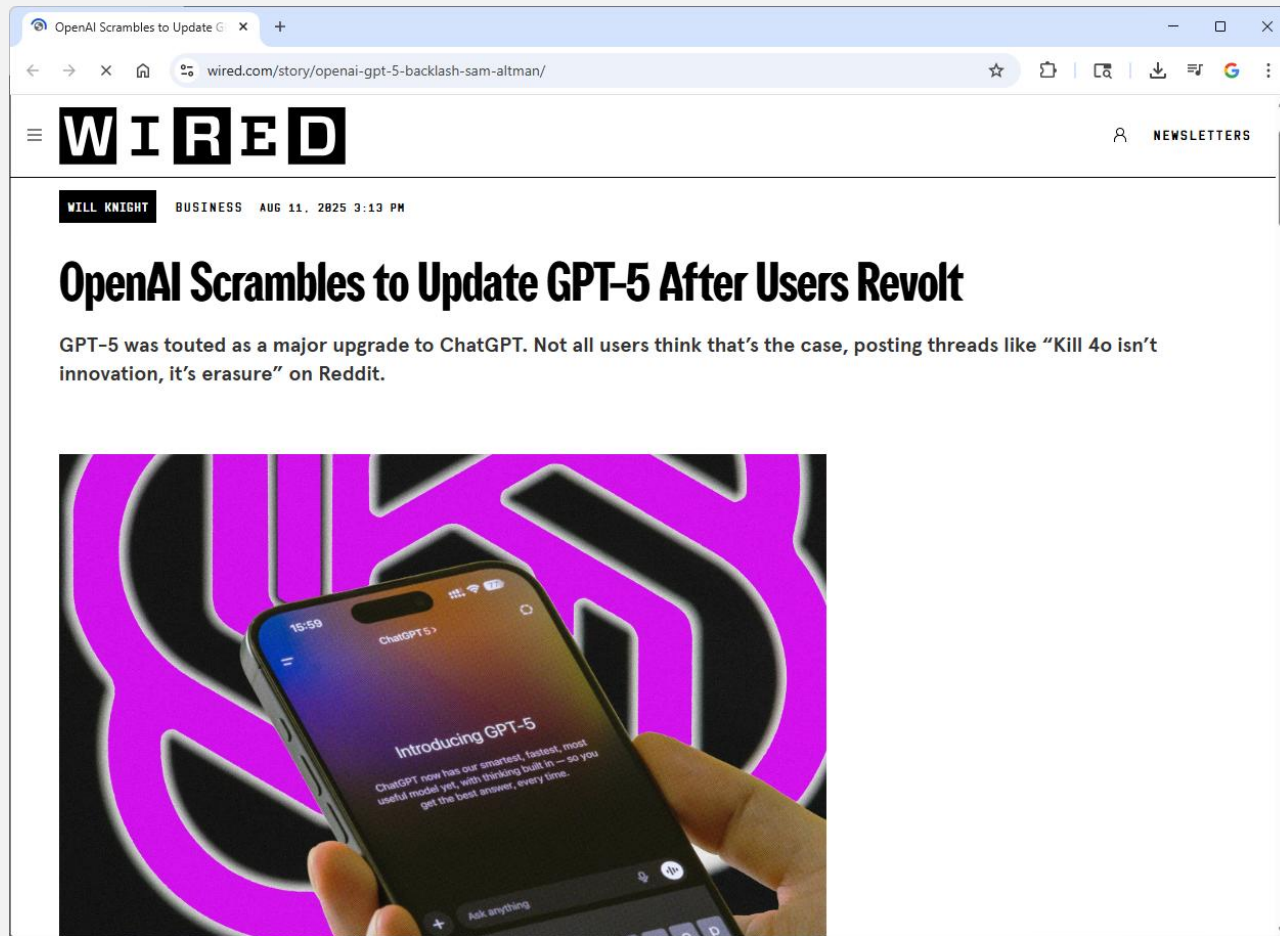
$$\frac{\text{Response Tokens}}{\text{Prompt Tokens} + \text{Response Tokens}} = \frac{1387}{22 + 1387} \quad \text{Downstream Fraction} = \mathbf{98.4\%}$$

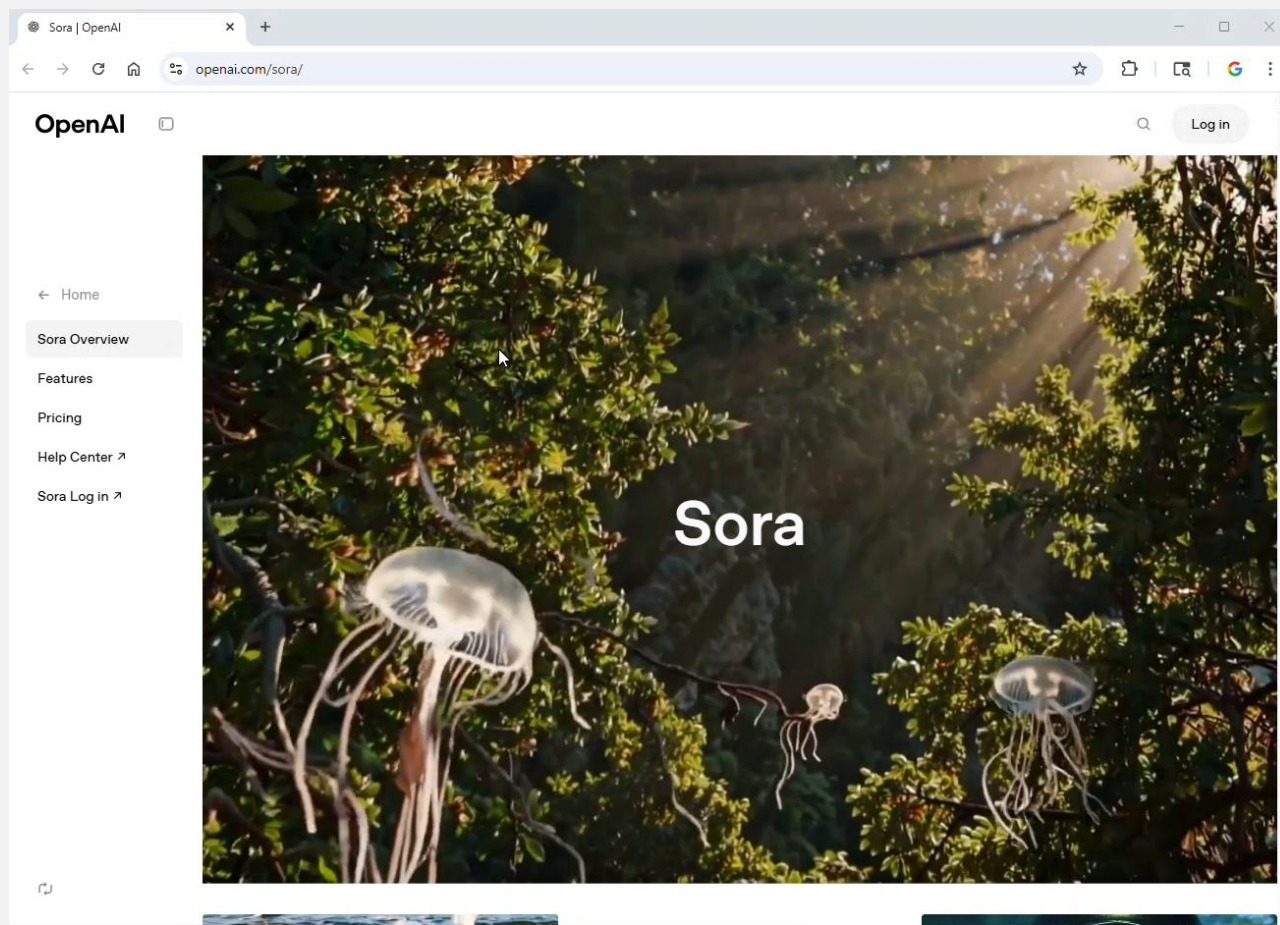
# AI Applications Upstream / Downstream Traffic Fraction



Source: Nokia







## Math time again

Sora Video Max Length = 20s

Generation Runtime (typ.) = 60s

$$\frac{\text{Generation Runtime}}{\text{Video Asset Duration}} = \text{Target Speedup Factor} = \mathbf{3!!}$$

ACE for Games | NVIDIA Developer

developer.nvidia.com/ace-for-games

NVIDIA DEVELOPER Home Blog Forums Docs Downloads Training

Topics Platforms and Tools Industries Resources

Home / ACE / ACE for Games


On this page

- Key Benefits
- Partner Experiences Powered by AI
- Get Started with NVIDIA ACE
- More Resources
- On-Demand Sessions

## NVIDIA ACE for Games

NVIDIA ACE is a suite of digital human technologies for middleware and game developers that powers knowledgeable, actionable and conversational game characters using generative AI. ACE provides ready-to-integrate cloud and on-device AI models for each aspect of digital humans—from speech to intelligence to animation.

[Get Started](#) [Notify Me](#)



The image shows a digital character of a chef, likely created using NVIDIA ACE technology. The character is an elderly man with a friendly expression, wearing a black chef's hat and a black short-sleeved shirt. He is standing in a kitchen, with a white clock on the wall behind him and various kitchen equipment visible in the background. The lighting is dramatic, highlighting the character's face and the details of his attire.

# NOKIA BELL LABS

“The dynamic element here is to represent a physical environment and the changes taking place in it. As the space changes, the representation changes with it...”

“...it’s like ChatGPT but for physical spaces. Instead of a large language model, this is a large world model.”

**Dr. Thierry Klein on Industry 5.0**



NOKIA