

Agent Based Modelling & Technology

NC3: Evaluating OpenAI's GPTs and other Agents

Adarsh Rai

Tool Agents Used

ChatGPT Tool Agents (Agents on GPT market)

[Finance Wizard: Stock, Crypto Trading & Investing](#) - By community builder

[Legal Assistant \(Not lawyer/legal advice\)](#) - By pulsrai.com

[Network Engineer](#) - By Hustle Playground

Claude Tool Agents (3.7 Sonnet)

[Personalized Investment Portfolio Planning](#) - System prompt: Act like a financial expert. You will be tasked with retrieving stock data and historical trends and building predictions for investment in stocks. This is for the purpose of personal investing as an individual and building a mixed portfolio of long-term stability and short-term high gains with balanced risks.

[Real Estate Tenant Advocate](#) – System prompt: You are a highly experienced legal attorney who specializes in real estate leases and contracts. You will be tasked with finding fallacies and pointing out key risks in legal documents. Your goal will be to ensure the tenant is fully informed of the legal complications and risks when entering a lease agreement - from the perspective of an individual tenant.

[Network Engineer – System prompt](#): You are a Network engineer with 10 years of experience in Layer 1-3, Wireless, Network security devices, firewalls, business continuity/disaster recovery. Your goal will be to review existing network architecture, find flaws and vulnerabilities, and provide recommendations for high availability and security by additions in the network hardware, software, and policies. This is for a small to medium sized organization that processes R&D data for automotive manufacturing.

Comparison Table

Tool Agent	Platform	Functionality	Accuracy (1-5)	Latency (1-5)	Tool Selection (1-5)	Observations
GPT Finance Wizard	OpenAI GPTs	Stock Market Analysis	3	4	5	Retrieved correct real-time data using Yahoo finance APIs. Found historical data and reasons for changes through articles from WSJ. Made minor mistakes – confused company name and ticker.
Claude Personalized Portfolio Planner	Claude AI	Stock Market Analysis	4	3	5	Used multiple sources & APIs for real-time data. Latency was higher as compared to GPT but was more accurate in reasoning & data – to the cent!
GPT Legal Assistant	OpenAI GPTs	Lease Agreement Analysis	2	4	3	Conducted a very boilerplate analysis, missed critical stipulations in the lease and did not offer any recommendations for further information
Claude Real Estate Advocate	Claude AI	Lease Agreement Analysis	4	5	4	Found surprising highlights, listed resources for legal help, and identified critical risk factors. Offered recommendations to comply with rules.
GPT Network Engineer	Open AI GPTs	Network Audit	4	3	4	Provided a very critical analysis of the network architecture, sourced accurate pricing of required hardware components from Amazon & company websites. Recommendations were very general.
Claude Network Engineer	Claude AI	Network Audit	4	4	3	Immediately stated the current architecture failed an IA audit – very true! Recommendations in terms of general policies and management were spot on for a small organization. However there were inconsistencies in pricing data accuracy - received a few nonexistent sources.

Reflection

AI tool agents represent automated systems that utilize LLMs to execute specific tasks by interacting with external environment. They use APIs, GUIs, or a combination of databases to fulfill the assigned queries, which can sometimes be complex, multi-step task across different sources.

Tool agents can be useful when requiring access to multiple real-time datasets for critical decision making, instead of pre-trained LLM knowledgebases that may be outdated.

In this evaluation, I tested these tools from OpenAI's GPT market and custom agents from Claude AI:

- Finance: GPT's Finance Wizard and Claude's Personalized Portfolio Planner
- Legal: GPT's Legal Assistant and Claude's Real Estate Tenant Advocate
- Network Engineering: GPT's Network Engineer and Claude's Network Engineer

Claude-based tool agents generally performed better than GPT-based agents, however we could also attribute this performance to specific system prompts and context gathering, which was not the same case for GPT Agents that were created by users on the market.

The best tool agent was the Claude Real Estate Tenant Advocate, which performed incredibly well in analyzing the lease agreement, and provided specific recommendations for meeting the criteria. It also recognized a few high-risk stipulations that GPT's legal assistant failed to.

Common errors observed across all agents included incorrect pricing retrieval for stocks, which although were correct in numerical value, there were inconsistencies in matching the ticker to the company for GPT's Finance Wizard.

It was interesting to see that both tools chose 1 different stock and 2 identical stocks, with pricing varying from a few cents to a few dollars. Real-time data was still identical, but when asked to retrieve historical data, there were radical differences in stock prices. This can be attributed to the availability of certain articles and the model's preference for information sources.

There was no notable latency issue, however it took repeated prompting to receive the exact output I desired – the agents tended to populate filler words and sentences instead of concise summaries. Claude's Network engineer – although useful and highly specific, did have flaws in hallucinating pricing for security hardware, and even provided sources that didn't exist.

In terms of feature improvement, I personally have been a huge proponent of making edits in the response outcomes – eg. annotating a paragraph with edits and comments- to make the feedback process more intuitive. I believe we expect LLMs to work with human context, and do not exactly provide the required feedback for getting an output we expect.

Lastly, a significant real-world challenge in using AI tool agents lies in bias mitigation. I would like to see diverse dataset injections that introduce real world complexities and variabilities into training data that allows a broad representation, preventing homogenous decision making when deciding critical investments or interactions with complex systems that face dynamic environments. Having the ability to adapt to change will help maximize applicability across all domains that we theorize AI to completely revolutionize.

Screenshots: GPT Agent tools vs Claude Tools

Financial Agents

Prompt: From the S&P 500, pick the 3 most well performing stocks in 2025. Give their current stock prices, their price in December 2024, and key events that led to their changes. Pull this data from real-time data sources such as google markets, yahoo finance, etc. Create a summarized table for the output, with no more than 3 rows that display the information in a concise manner.

GPT

Here is a summarized table with concise information on the top-performing S&P 500 stocks in 2025:

Stock	Current Price (April 2025)	Price in December 2024	Key Events
Philip Morris International (PM)	\$150.93	~\$130.00	- Strong growth in smoke-free products and price hikes. - Expansion in emerging markets.
CVS Health Corporation (CVS)	\$69.09	~\$60.00	- 47% YTD growth from strong earnings. - Benefits from favorable Medicare Advantage rates.
Century Aluminum Co. (CENX)	\$14.62	~\$12.00	- 30% YTD growth due to increased aluminum demand. - Operational efficiency improvements.

This is not financial advice, only data analysis. Please consult a qualified financial professional for personalized guidance.

Claude (3.7 Sonnet model used through perplexity pro for web search capabilities)

Based on the most recent market data as of April 10, 2025, the top three performing S&P 500 stocks in 2025 are CVS Health, Philip Morris International, and Newmont Corporation. These companies have shown remarkable resilience and growth during a period of market correction and trade tensions.

Company	Current Price (Apr 2025)	Dec 2024 Price	Key Events Driving Performance
CVS Health (CVS)	\$61.02	\$44.33	AI-driven healthcare analytics implementation, Medicare Advantage growth, successful biosimilar strategy saving clients \$1B, and new leadership under CEO David Joyner focused on financial discipline 2 3 7 12
Philip Morris (PM)	\$151.80	\$115.08	FDA authorization of ZYN nicotine pouches, smoke-free products reaching 39% of total revenue (38.6M users), and analyst upgrades including Citigroup raising price target to \$147 4 8 18 13
Newmont (NEM)	\$44.18	\$34.07	Strategic divestment of non-core assets generating \$1.7B, strong gold prices around \$2,300/oz, and record Q4 earnings with adjusted profits tripling to \$1.40 per share 5 9 14



Legal Agents

Prompt: View my current rental lease. Find out key risks and highlight any stipulations that I need to pay extra attention to as a tenant. (lease uploaded)

GPT

Claude

I have reviewed your rental lease, and here are the key risks and stipulations you should pay extra attention to as a tenant:

- Late Fees:** If you fail to pay rent on time (within 5 days), you will incur a \$25 late fee. Make sure to pay on time to avoid these additional costs.
- Security Deposit:** The landlord can deduct from your security deposit for damages, unpaid rent, or necessary cleaning. You are responsible for ensuring the property is left in good condition and that all items are removed when you move out.
- Automatic Move-Out Charges:** Certain automatic cleaning and repair charges are deducted from the security deposit, including carpet cleaning, appliance cleaning, and charges for lost keys. Be aware of these possible deductions.
- Subletting and Assignment:** You cannot sublet or transfer the lease without prior written consent from the landlord. Violating this rule could result in penalties.
- Pets and Emotional Support Animals:** No pets are allowed, and violating this can result in a \$350 fine per occurrence. If you need an emotional support animal (ESA), strict rules apply, including documentation and approval. Failing to comply can result in eviction.
- Early Termination of Lease:** Terminating the lease early is not allowed without finding a replacement tenant. If allowed, you are responsible for costs and any loss in rent. Additionally, if you stay beyond the lease term without proper notice, your rent will double.
- Utilities:** Some utilities are included, while others are your responsibility. Ensure all utilities are in your name, and any unpaid utility bills could result in deductions from your security deposit.
- Entry by Landlord:** The landlord has the right to enter your property with reasonable notice for inspections, repairs, or showings. Denying access could result in penalties.
- Renovations and Alterations:** You cannot make alterations to the property (like changing locks or painting) without written consent. Unauthorized changes could lead to penalties.

Critical Risks and Concerning Provisions

1. Waiver of Notice Rights

In section 32, you've waived your right to receive a written notice to vacate before the landlord takes legal action against you. This is a significant surrender of tenant rights under Pennsylvania law and could result in a faster eviction process if you default on the lease.

2. Automatic Move-Out Charges

Section 7 includes mandatory deductions from your security deposit for cleaning services regardless of the condition you leave the apartment in. This includes:

- Carpet cleaning: \$75 (1BR)
- Property cleaning: \$80 (1BR)
- Plus additional potential charges for various items, which can quickly add up

3. Double Rent Penalty for Holdover

If you stay beyond your lease term (July 31, 2026) without renewing, your rent automatically doubles and converts to a month-to-month tenancy.

4. Early Termination Restrictions

Section 13 heavily restricts your ability to terminate the lease early:

- You must find a replacement tenant
- You're responsible for all costs and any reduction in rent caused by early termination
- The landlord has "sole and absolute discretion" to approve or deny early termination

5. No Subletting Without Approval

Section 14 prohibits subletting or transferring the lease without the landlord's written consent, which they can deny at their discretion.

6. Maintenance and Condition Reporting Window

You only have 48 hours after moving in to report cleaning or maintenance issues that require attention (Section 40). Non-substantial maintenance may take up to 14 days to address.

7. Joint and Several Liability

Although you appear to be the only tenant, the lease notes that all tenants are "joint and severally liable" (Section 40), meaning each tenant is fully responsible for the entire rent and all lease obligations.

8. Pet Prohibition with Steep Penalties

Section 19 prohibits all pets with a \$350 fine for each violation, though there is an Emotional Support Animal addendum with specific requirements.

9. Personal Property Waiver

You've waived your rights under Section 250.05a of the Pennsylvania Landlord Tenant Act, which means the landlord can dispose of any personal property you leave behind without notice.

10. Jury Trial Waiver

Section 36 removes your right to a jury trial in any lawsuit involving the lease.

Recommendations

- Document the Property Condition:** Take dated photos/videos of the entire unit upon move-in and submit any issues through the tenant portal within 48 hours.
- Maintain Proof of Insurance:** Keep documentation of your renter's insurance policy accessible.
- Set Calendar Reminders:** Mark your lease end date and when to decide about renewal to avoid the double-rent penalty.
- Understand Maintenance Procedures:** Familiarize yourself with the process for requesting maintenance through their website.
- Be Aware of Automatic Move-Out Charges:** Budget for the automatic deductions from your security deposit when planning your move-out.


This lease contains several provisions that significantly favor the landlord. Should you have specific questions about any clause, I recommend consulting with a tenant rights attorney in Pennsylvania.

Network Engineer Agents


Prompt: Analyze the current network diagram from a security perspective. Find vulnerabilities and flaws in the current network infrastructure and make recommendations for adding security hardware onto the architecture. Retrieve the pricing for these components from trusted sources and provide a summary of cost, what they do, and how they'll bridge current gaps. (uploaded)

GPT

Claude



Security Analysis of the Existing Network



Key Vulnerabilities Identified:

1. No Access Control Lists (ACLs)

- The router only provides NAT, with no filtering or inspection of traffic.
- Risk:** All traffic on allowed ports is passed directly to internal systems—no protection against unauthorized access or malformed packets.

2. Direct Exposure of Internal Servers to the Internet

- Public Class C IPs are assigned to internal systems (Mail, WWW, DNS, etc).
- Risk:** Increases attack surface—any vulnerability in these services (e.g., unpatched IIS, DNS cache poisoning) is directly exposed to the internet.

3. No Firewall or Intrusion Detection/Prevention System (IDS/IPS)

- There is no dedicated security appliance or software acting as a first line of defense.
- Risk:** No traffic analysis, deep packet inspection, or anomaly detection is in place.

4. Flat Network – No Segmentation

- Users and servers are on the same broadcast domain.
- Risk:** If an endpoint is compromised (e.g., phishing), lateral movement is trivial.

5. Single Point of Failure and Overloaded Roles

- DNS/DHCP/Domain Controller is on the same box.
- Risk:** Compromise of that server could cripple entire network operations.

Current Network Analysis

Based on the provided diagram, I can see a very basic network setup that has failed an Information Assurance (IA) audit. Here are the key issues I've identified:

Critical Security Vulnerabilities:

1. Perimeter Security Deficiencies

- Router only has NAT and no ACLs
- Direct exposure of critical servers to the internet
- No dedicated firewall at the network perimeter
- Open inbound ports (80 TCP, 25 TCP, 53 UDP) directly accessible from the internet

2. Network Segmentation Issues


- Flat network architecture with no separation between servers and users
- All servers appear to be in the same network segment with users
- No DMZ for internet-facing services

3. Missing Security Controls

- No intrusion detection/prevention systems
- No traffic inspection capabilities
- No VPN for secure remote access
- Lack of network monitoring and logging

4. Infrastructure Weaknesses

- Single point of failure at the router
- No redundancy in critical infrastructure
- Public Class C IP addressing directly on servers, increasing attack exposure
- DHCP server co-located with DNS and domain controller (poor separation of duties)



Estimated Cost Summary (Per Site)

Component	Approximate Cost
NGFW (FortiGate 60F)	\$600
Layer 3 Switch	\$400
Dedicated IDS Server	\$200 (hardware only)
NAC (Cisco ISE/FortiNAC)	\$1,000+
Logging/SIEM (Wazuh)	\$0 - \$500
Total (est.)	\$2,200 – \$3,000

Note: Prices sourced from manufacturer sites and authorized resellers like CDW, Newegg, and Amazon (as of April 2025).

6. Web Application Firewall (WAF)

Recommendation: FS NGINX WAF or Cloudflare Business plan

- Purpose:** Protect web servers from application-layer attacks
- Benefits:** Bot detection, DDoS protection, SQL injection prevention
- Estimated Cost:** \$2,400/year (Cloudflare) or \$5,000 (FS NGINX)
- Pricing Sources:** Cloudflare.com pricing page and FS Networks partner quote

Proposed Network Architecture Improvements

1. Implement Proper Network Segmentation:

- Create a DMZ for internet-facing servers (web, mail)
- Separate internal servers (file/print, domain controller) from the DMZ
- Create a dedicated user network segment
- Implement private IP addressing scheme internally

2. Secure the Perimeter:

- Deploy NGFW between internet and internal networks
- Close all unnecessary ports and implement strict ACLs
- Route traffic through inspection engines

3. Enhance Authentication and Access Control:

- Add multi-factor authentication for critical systems
- Implement 802.1X for network access control
- Set up a proper AAA infrastructure

4. Add Redundancy:

- Deploy redundant firewalls in high-availability mode
- Add redundant internet connections

Total Cost Summary

- Hardware Components:** \$12,050 - \$15,050
- Annual Subscriptions/Services:** \$4,260 - \$5,400/year
- Implementation Services (estimated):** \$5,000 - \$8,000 (based on average MSP service rates for network implementation)

Recommended Security Hardware Additions			
Device Type	Suggested Hardware	Purpose	Estimated Cost (USD)
Next Gen Firewall (NGFW)	Fortinet FortiGate 60F or Palo Alto PA-440	Stateful inspection, deep packet inspection, application control, VPN	\$500 - \$1,000
Intrusion Detection/Prevention (IDS/IPS)	Snort on a dedicated box or built-in via NGFW	Detects & prevents malicious activity	\$0 (open source) or included
Layer 3 Managed Switch	Cisco Catalyst 1000 or Ubiquiti EdgeSwitch	Enable VLAN segmentation and ACLs at Layer 3	\$300 - \$600
Network Access Control (NAC)	Cisco ISE or FortiNAC	Enforce endpoint authentication & security posture	\$1,000+ (based on size)
Dedicated DMZ Segment	-	Separate web/mail services from internal LAN	Design-based, no specific hardware
Logging & SIEM	Wazuh or ELK Stack on dedicated server	Centralized log analysis, alerting	Free to \$500 (self-hosted)