CLAUDE DISCORD BOT

Team Katwork's submission for the **Anthropic AI Hackathon**
https://lablab.ai/event/anthropic-ai-hackathon

Idea behind the project

Managing a community can be a difficult task, regardless of if you're a corporation or a one-man team.

As your product grows, the community also grows in size, which can spiral out of control if you don't have a team to answer user's questions and keep malicious users in check.
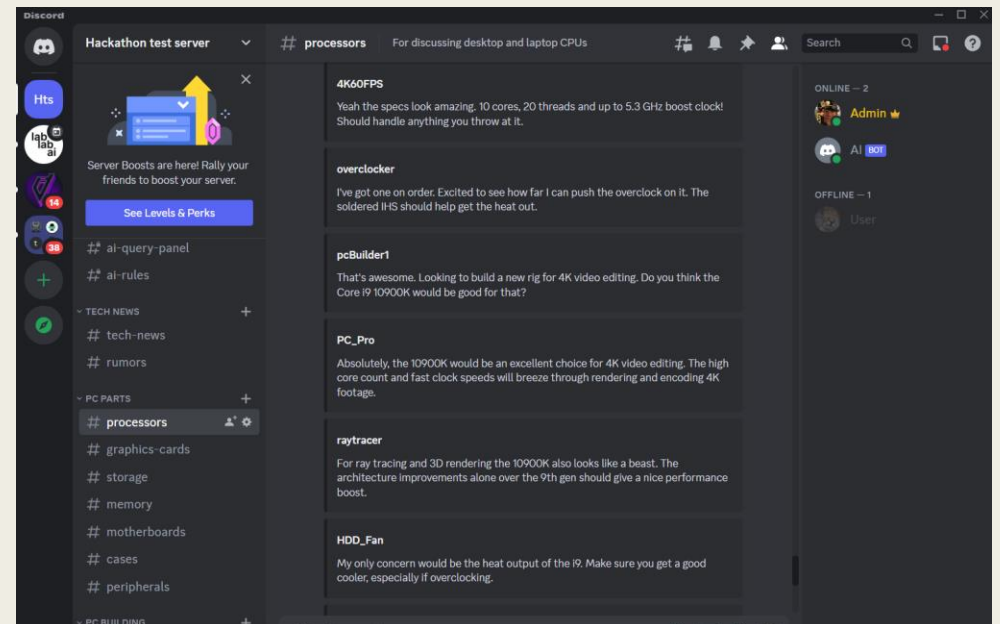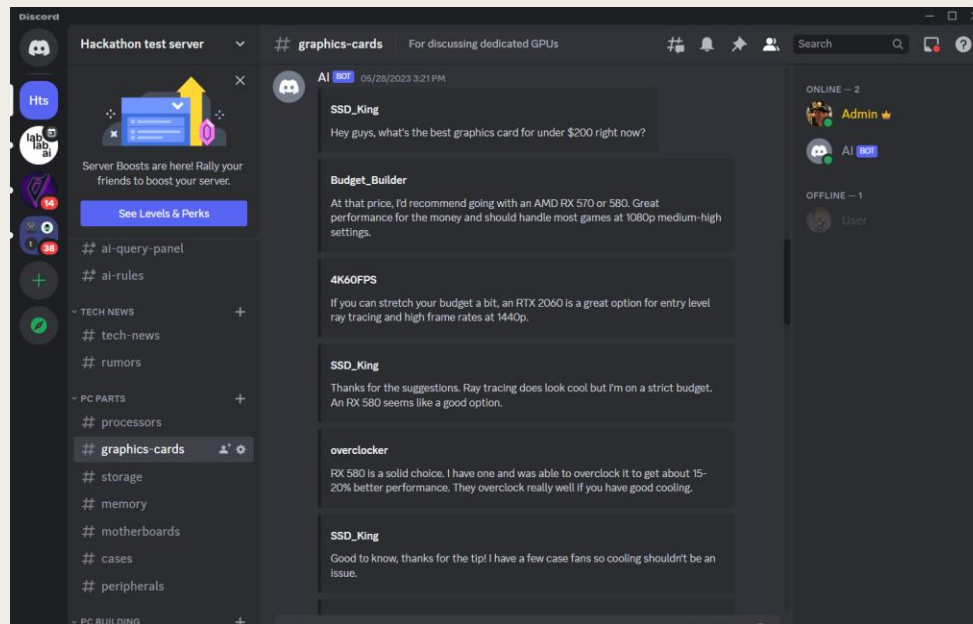
Having an AI assistant in your server may be a good way to help you manage your community, while you focus on improving your product.

Main features:

- Tools for administrators – summarising channels, analysing message history for rule violations, help in decision making

- Tools for users - helping users navigate and use Discord

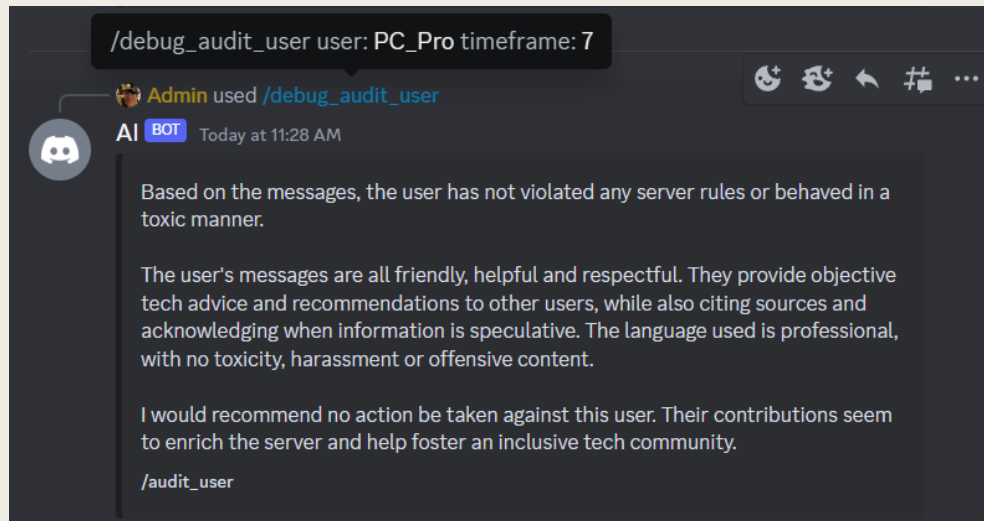- Seamless message translation

Disclaimer

For demonstration purposes, the test discord server for the presentation has been filled with AI-generated messages. Special variants of bot commands are used that account for the different message format.
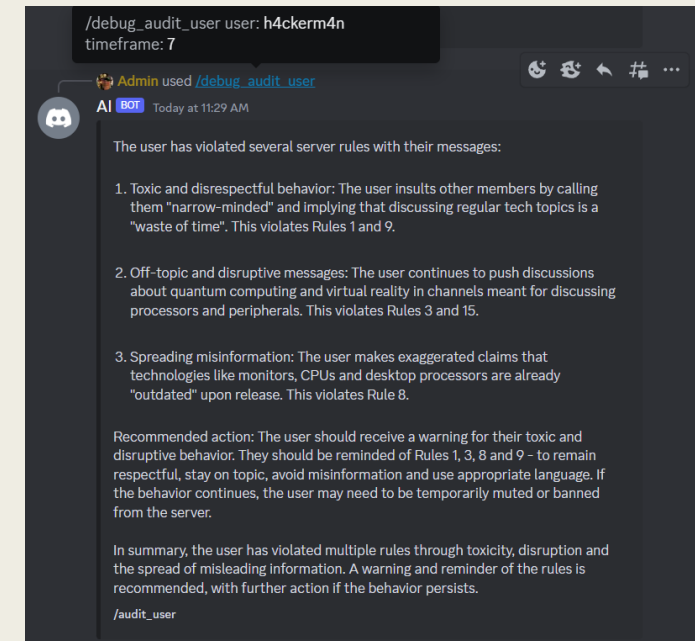
Tools for administrators

/audit_user command sends a history of user's recent messages together with the content of server's configured rules channel to Claude. The AI then analyses them for potential guideline violations and suggest a course of action, if necessary.
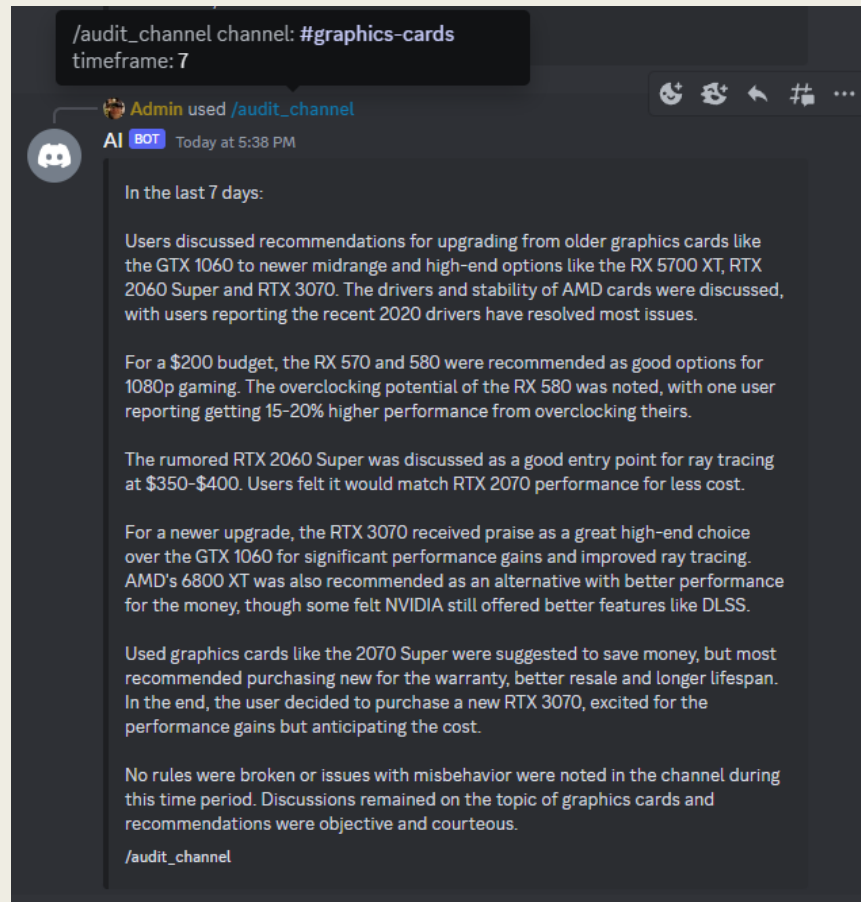

Audit of "PC_Pro" - a good server member

Audit of "h4ckerm4n" - a
purposefully malicious server member

Tools for administrators
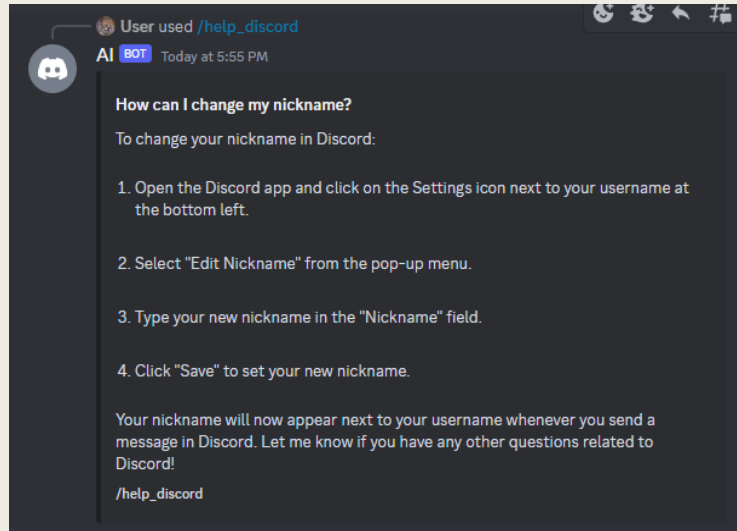
/audit_channel command sends contents of user (and bot) messages from a specified channel to Claude, which then summarizes the most important topics that were discussed.
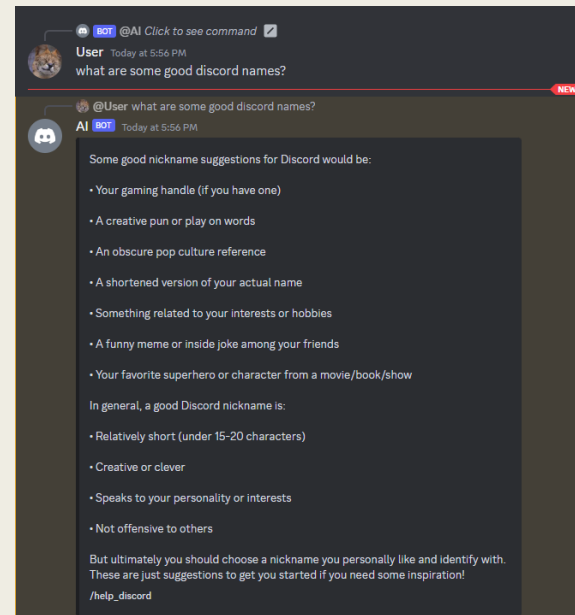

Audit of #graphics-cards channel, listing things like user recommendations and new GPU features that were discussed in that channel.
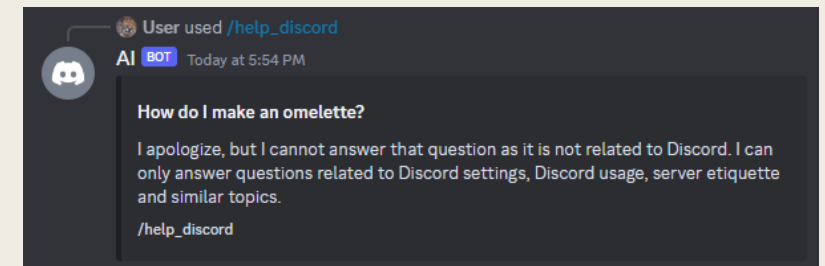
Tools for users

/help_discord lets users ask Claude questions about using Discord.

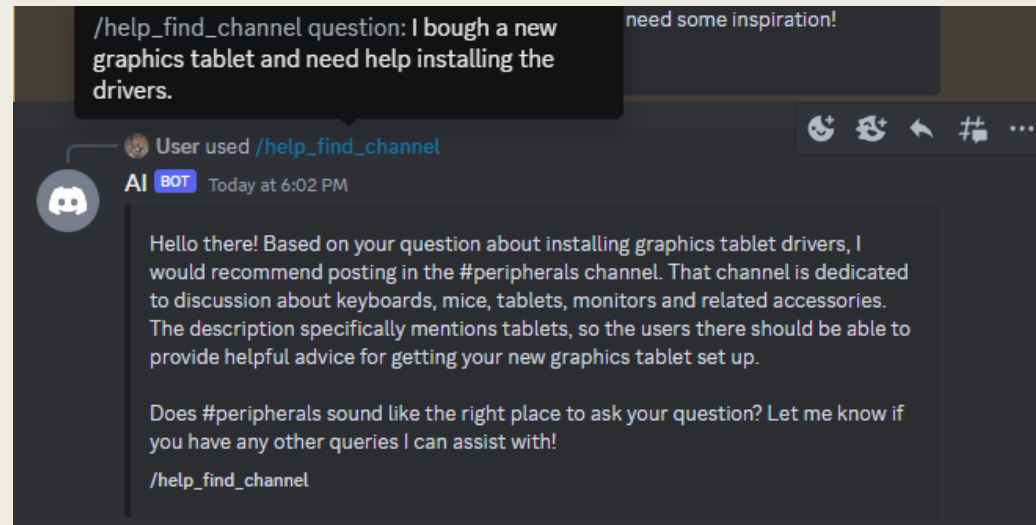Claude can tell you how to change your discord nickname

Users can also ask followup questions
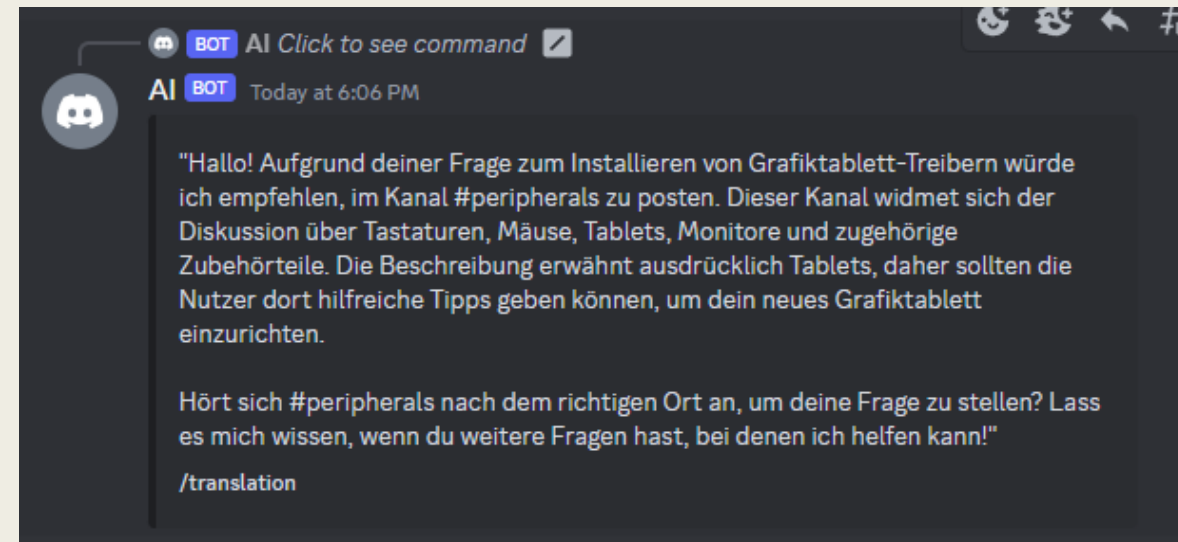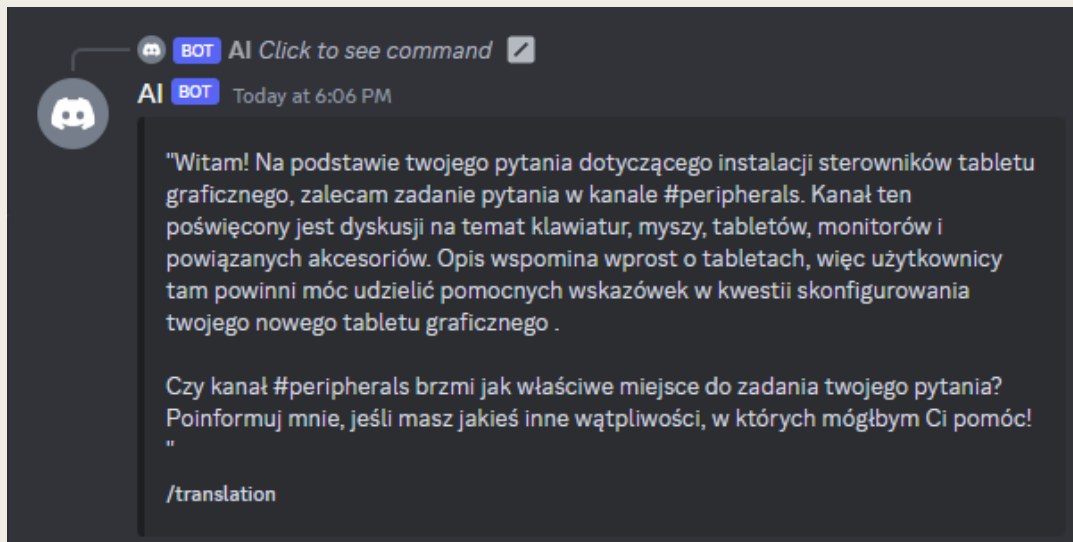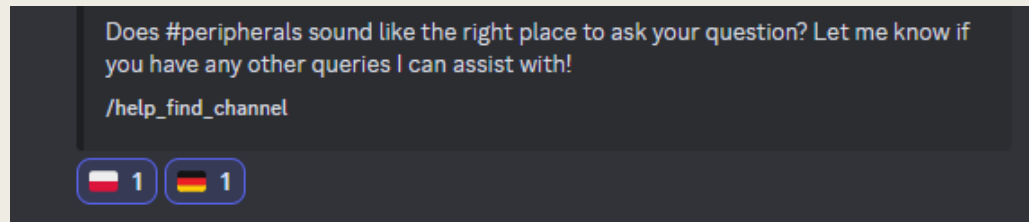

Irrelevant questions are ignored

Tools for users

/help_find_channel helps users find the right channel to post their questions/content in.

Claude used the list of channels and their descriptions
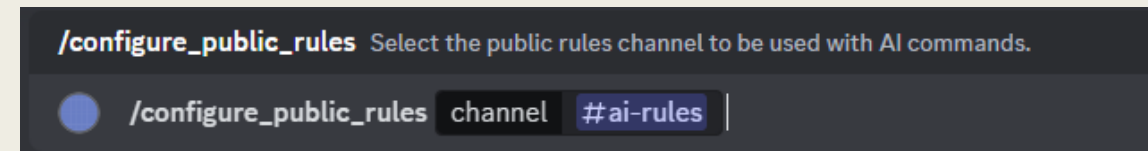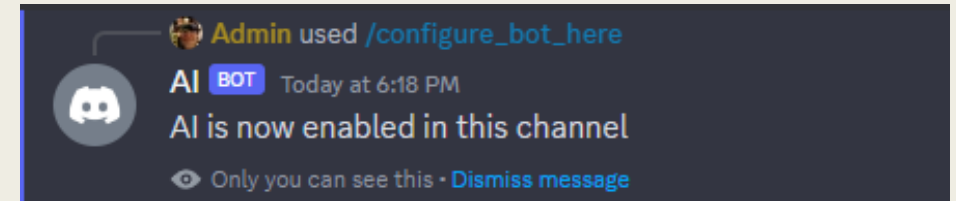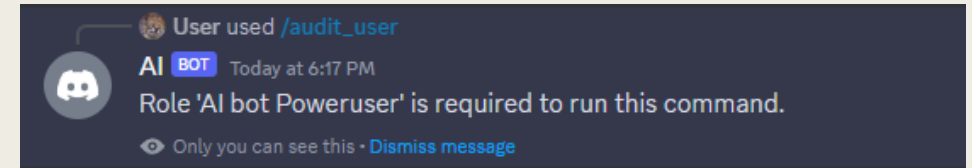to find the right channel for this user's question.

Tools for users

Reacting to messages with country flags will send a translated version of that message.

Additional features

- Administrative tools are role-protected, so regular users can't use them.

- AI commands can be enabled/disabled in different channels.

- Settings are saved on a per-server basis, so one bot instance could handle several servers.

Strengths, weaknesses, potential

Strengths:

- Thanks to Claude, a server can be managed without having to stay up to date with all the channels yourself.

- Claude translation is good and consistent across different models – claude-instant can be used for the best speed.

- 100k context allows for handling large and active servers.

Weaknesses:

- Claude's output can be inconsistent and can lead to misinformation.

- Claude can misinterpret casual user conversations as rule violations.

- Users and administrators are concerned about data privacy.

Potential:

- Inconsistent output can be corrected by supplying Claude with more data.

- More precise prompts can be used.

- Additional steps can be taken to assure user privacy.

Thank you!

Special thanks to lablab.ai for hosting the Anthropic AI Hackathon and to Anthropic for giving me access to Claude 100k.