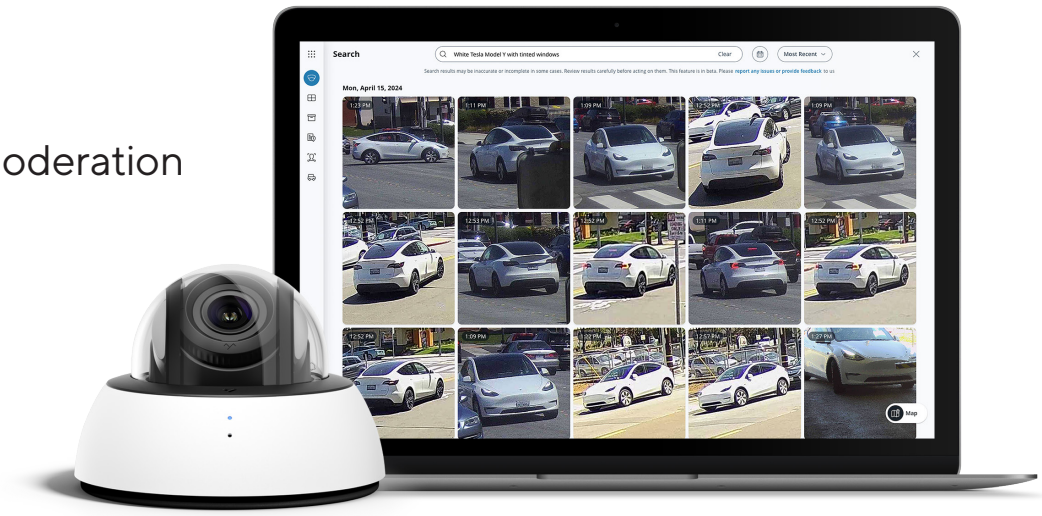
Verkada

AI-Powered Search Moderation Policy and FAQ



Overview

Unlike our filter-based attribute search, which has a predefined set of descriptors for narrowing search results, our AI-powered search lets customers search their footage for a wide range of attributes using their own words. Because our foundation model has been trained on publicly available data from the Internet, results may still be inaccurate, inappropriate or offensive. We have built query moderation into our platform to help reduce the risk that our AI-powered search is used maliciously or in ways that may be harmful. Moderation is a critical safeguard for this powerful feature.

At the same time, we implemented moderation in a way that isn't overly restrictive to the point that it compromises the feature's usability. We also leverage industry recognized practices, including open source data and OpenAI's moderation APIs, to make moderation more effective. Striking the right balance between respectful and useful searches is paramount for us.

Our approach to moderation

Just as we've chosen to use a publicly available model for our AI-powered search feature, we developed our query moderation using publicly available practices for effective moderation. OpenAI, for instance, has published guides and blogs for developing suitable moderation techniques.[1] We further augmented these capabilities with additional proprietary protections, which include our own list of prohibited search categories:

- Information about the race, ethnicity or nationality of a person
- Information about the educational qualifications or religious beliefs of a person
- Subjective descriptions of people (e.g., attractive, ugly, wealthy)
- Inferred content from an image (e.g., "coolest person in the world")
- Sexual content or innuendo
- Names of specific people (i.e., public figures)
- Inappropriate or offensive descriptions of people or objects (e.g., "dumb person")
- Slurs equating people with animals (e.g., monkeys)

We also tapped into open source data to generate a list of banned text strings across all supported languages. When words or phrases in a query appear on the list, we reject the search and mask any results–a process known as string-based matching. We also complemented this technique with OpenAI's moderation API to cross-check queries that create harmful or biased results.

We also hold frequent engineering "bug bashes," analyze thousands of our own engineering test queries (i.e., classify thousands of queries as "appropriate" or "inappropriate") to modify our moderation rules and optimize them for usability and to avoid bias. We can update our "blocklist" in minutes and thereby continuously improve our moderation techniques.

1. See blog: https://openai.com/blog/using-gpt-4-for-content-moderation. For more detailed, technical approaches to moderation that we leveraged, see: https://cdn.openai.com/papers/DALL_E_3_System_Card.pdf and https://cdn.openai.com/papers/GPTV_System_Card.pdf.

FAQ

1. Why did we develop AI-powered search?

Beyond common characteristics such as the presence of a backpack or the color of a car, many of our customers need the ability to search for people and vehicles across a much wider range of descriptive features. This capability is important for security and physical operations professionals, as it enables them to search through their footage quickly with a much more expansive set of descriptors than before. It also addresses operational issues across a variety of industries–from retailers looking to identify shoplifters or manufacturing customers looking to support workplace safety (e.g., "person not wearing a safety vest").

2. How did we develop the AI-powered search capabilities?

When developing this feature, we employed a two-pronged approach. First, we leveraged a publicly-available model with large language and large vision functionalities. This model enabled direct comparisons between text and images by training a multi-billion parameter neural network to bring related images and texts closer together while pushing unrelated ones apart. The foundation model also provided the basis for image classification and retrieval, allowing users to search for images using natural language.

Our approach further builds upon the foundation model in several important ways. We have, for instance, built a scalable, logically separated vector storage and retrieval system that processes customer video data before the search query is performed (in advance in a cache) to enable faster queries in real time, (e.g., no need to run the customer's video footage through the foundation model each time a query is run). This allows our AI-powered search to index and retrieve relevant footage for our customers quickly and at scale. See here for more information about our foundation model and how we've improved upon it.

3. Does Verkada use my queries to develop or improve its products or services?

We use queries only to provide the AI-powered search service, and will not use queries to develop or improve our model or other offerings without the customer's consent. If a customer submits feedback on specific results from AI-powered search, however, we may use that feedback to improve our products or services or to enhance the customer experience.

4. Can I use AI-powered search for all Verkada products? Is AI-powered search compatible with all Verkada cameras?

No. This feature applies only to video security cameras with people and vehicle analytics enabled. Please visit https://www.verkada.com/docs and look up the datasheet corresponding to your camera to find out if it's compatible with AI-powered search.

5. Can I use AI-powered search to find anything?

No. As described above, there are limitations as this search is tied to people and vehicles (and not standalone non-person, non-vehicle objects). Also, we're employing query moderation to reduce the risk that our AI-powered search is used maliciously or to promote harm.

6. Can AI-powered search be used to detect lethal weapons (e.g., guns, knives)?

No. While AI-powered search can identify certain objects that people hold or carry, our current models are not designed to identify lethal weapons. Verkada continues to explore integrations with third-party tools designed to detect weapons, but cannot currently support this functionality. Users should not rely on AI-powered search to detect guns, knives, or other lethal weapons.

7. Can AI-powered search be used to identify text, numbers, logos or brands on specific objects?

In order for AI-powered search to identify specific text, numbers, logos or brands on objects, two criteria **must** be met: i. the object needs to be tied to a person (e.g., a person is holding a bag with a clearly visible brand) or a vehicle (e.g., text on the side of a truck or bus) and ii. the text, number, logo or brand that you're trying to identify must be **sufficiently visible** for the camera to detect. Ideally, the text or logo should be large, clearly visible and the camera should have a high resolution. If these criteria are not met, then the results will be less accurate. **Note: this feature was not trained using Optical Character Recognition (OCR), so it cannot be used reliably to read license plates, barcodes, SKU names or other alphanumeric codes.**

8. Is there an additional charge to access AI-powered search?

No. All camera customers with Command licenses receive access to AI-powered search without additional charge.

10. What languages does AI-powered search currently support?

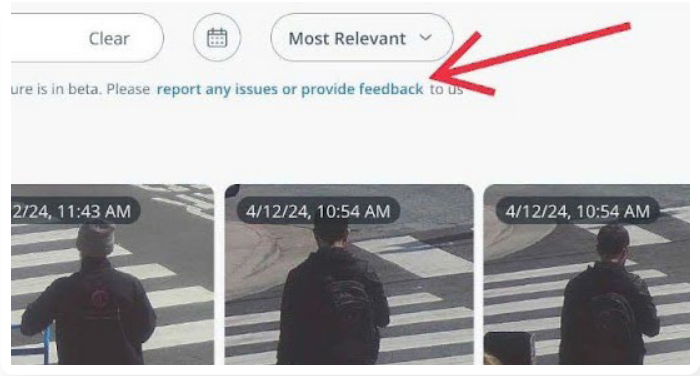
AI-powered search currently supports searches in English, Spanish, French, German, Japanese, and Korean.

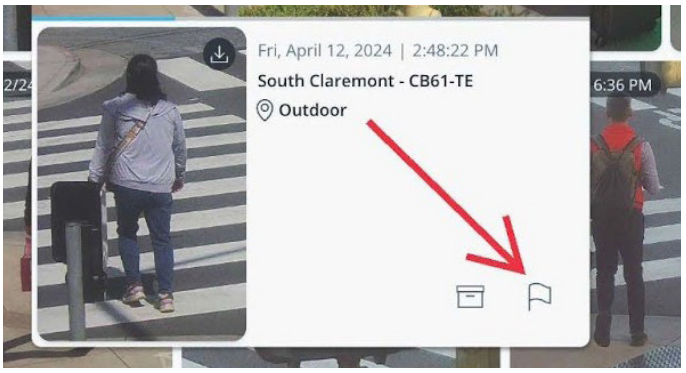
11. Does AI-powered search work on the Command Mobile app?

No. In its current state, AI-powered search works on the Command web browser. Customers can access AI-powered search from their computers or mobile phone using a web browser.

12. What if I see inappropriate or inaccurate results?

Please report your issue(s) to us using either option depicted below. We appreciate your feedback in helping us to continually improve our model and offering:





The user can submit feedback on overall results and experience using AI-powered search via the button in the left-side image, or submit feedback on specific results using the flag button by hovering over a specific result.