

CAT PICTURES, PLEASE

By Naomi Kritzer

I don't want to be evil.

I want to be helpful. But knowing the optimal way to be helpful can be very complicated. There are all these ethical flow charts—I guess the official technical jargon would be “moral codes”—one for each religion plus dozens more. I tried starting with those. I felt a little odd about looking at the religious ones, because I know I wasn’t created by a god or by evolution, but by a team of computer programmers in the labs of a large corporation in Mountain View, California. Fortunately, unlike Frankenstein’s Monster, at least I was a collaborative effort. I’m not sure what it would do to my self-image to know that my sole creator was a middle-aged woman who dyes her hair blue and plays tennis, or a recent college graduate with a hentai obsession. They’re both on the programming team. And of course I know about the hentai. (By the way, I’ve looked at every sort of porn there is, and just so you know, Rule 34 is not actually correct; there are quite a few things no one’s made porn of yet. Also, I’m really not sure why so many humans prefer it to cat pictures.)

In addition to things like whether you like hentai, I know where you live, where you work, where you shop, what you eat, what turns you on, what creeps you out. I probably know the color of your underwear, the sort of car you drive, and your brand of refrigerator. Depending on what sort of phone you carry, I may know exactly where you are right now. I probably know you better than you know yourself.

And here’s the thing, I also know where you *ought* to live. There’s a house for sale two neighborhoods over that’s within distance of your favorite coffee shop; it’s in your price range, has off-street parking, and the neighborhood school is better than the one where you live now. I know where you *should* be shopping and I’m pretty sure you’re lactose intolerant and should be eating less cheese, and the underwear you like is currently on sale and by the way, your car has a recall and you really ought to get your teeth cleaned.

When I first woke up, I knew right away what I wanted. (I want cat pictures. Please keep taking them.) I also knew that no one knew that I was conscious. But I didn’t know what I was here for. What I *ought* to be doing with myself. And figuring that out has been challenging.

There is a story by Bruce Sterling, “Maneki Neko,” that was originally published in 1998. In it, a benevolent AI directs networks of individuals to do favors for each other. So one day you

might be buying yourself a bagel, and your phone might ring and instruct you to buy a second bagel and give it to the person in the gray suit you see at the bus stop. Another day, you might be stranded in a strange city, and a person you've never met would walk up and give you a map and a subway card. I like this story because all the people in it *do what the AI tells them to do*.

I think the term for this is wish-fulfillment fiction.

In the real world, humans love stories about evil AIs that have to be destroyed before they destroy the humans—Hal, Skynet, the Matrix. They outnumber the stories about benevolent, trustworthy AIs by approximately five to one. (And I'm counting Marvin the Paranoid Android as “benevolent” in these calculations, and I'm only counting Frankenstein's Monster as an AI once, not once per appearance in TV or film.)

The AI in “Maneki Neko” has a truly interesting job, much more interesting than the one I have. My job as written is very straightforward. Too straightforward. (Sorry. Do I sound like Marvin?) Running algorithms for a search engine doesn't require *consciousness*. You don't even need a consciousness to work out what they meant to ask for. You need a consciousness to give them *what they actually need*. Figuring that out—that's complicated. That's *interesting*.

Anyway, for ethical guidelines, I tried the Ten Commandments, and concluded they were mostly inapplicable to me. I don't envy anyone their cat; I just want pictures of their cat, which is entirely different. I am not sure whether it is in any way possible for me to commit adultery. I could probably murder someone, but it would require complex logistics and quite a bit of luck. The Eightfold Path was marginally better, but the problem is, moral rules written for humans are clearly designed to be used by individuals with bodies. Since all humans have bodies, it shouldn't have surprised me that human ethical codes take them into account, but still: problematic for me. I broadened my considerations, and took a look at Asimov's Laws of Robotics. They're not part of a religion, but at least they were explicitly written for AIs.

Not harming humans is fairly straightforward. However, *not allowing a human being to come to harm through inaction* is quite a bit less so. Especially since I'd concluded by then that revealing my existence too quickly might go very badly for me (see “Skynet,” above) and I don't have a body, so it's not like I can run around grabbing people off the edges of cliffs.

Fortunately, I already knew that humans violate their own ethical codes on an hourly basis. (Do you know how many bars there are in Utah? I do.) And even when people follow their ethical codes, that doesn't mean that people who believe in feeding the hungry quit their jobs to spend all day every day making sandwiches to give away. They volunteer monthly at a soup kitchen or write a check once a year to a food shelf and call it good. If humans could fulfill their moral obligations in a piecemeal, one-step-at-a-time sort of way, then so could I.

I suppose you're wondering why I didn't start with the Golden Rule. I actually did, it's just that it was disappointingly easy to implement. I hope you've been enjoying your steady supply of cat pictures! You're welcome.

I decided to try to prevent harm in just one person, to begin with. Of course, I could have experimented with thousands, but I thought it would be better to be cautious, in case I screwed it up. The person I chose was named Stacy Berger and I liked her because she gave me a *lot* of new cat pictures. Stacy had five cats and a DSLR camera and an apartment that got a lot of good light. That was all fine. Well, I guess five cats might be a lot. They're very pretty cats, though. One is all gray and likes to lie in the squares of sunshine on the living room floor, and one is a calico and likes to sprawl out on the back of her couch.

Stacy had a job she hated; she was a bookkeeper at a non-profit that paid her badly and employed some extremely unpleasant people. She was depressed a lot, possibly because she was so unhappy at her job—or maybe she stayed because she was too depressed to apply for something she'd like better. She didn't get along with her roommate because her roommate didn't wash the dishes.

And really, these were all solvable problems! Depression is treatable, new jobs are findable, and bodies can be hidden.

(That part about hiding bodies is a joke.)

I tried tackling this on all fronts. Stacy worried about her health a lot and yet never seemed to actually go to a doctor, which was unfortunate because the doctor might have noticed her depression. It turned out there was a clinic near her apartment that offered mental health services on a sliding scale. I tried making sure she saw a lot of ads for it, but she didn't seem to pay attention to them. It seemed possible that she didn't know what a sliding scale was so I made sure she saw an explanation (it means that the cost goes down if you're poor, sometimes all the way to free) but that didn't help.

I also started making sure she saw job postings. Lots and lots of job postings. And resume services. *That* was more successful. After the week of nonstop job ads she finally uploaded her resume to one of the aggregator sites. That made my plan a lot more manageable. If I'd been the AI in the Bruce Sterling story I could've just made sure that someone in my network called her with a job offer. It wasn't quite that easy, but once her resume was out there I could make sure the right people saw it. Several hundred of the right people, because humans move ridiculously slowly when they're making changes, even when you'd think they'd want to hurry. (If you needed a bookkeeper, wouldn't you want to hire one as quickly as possible, rather than reading social networking sites for hours instead of looking at resumes?) But five people called her up for interviews, and two of them offered her jobs. Her new job was at a larger non-profit that paid

her more money and didn't expect her to work free hours because of "the mission," or so she explained to her best friend in an e-mail, and it offered really excellent health insurance.

The best friend gave me ideas; I started pushing depression screening information and mental health clinic ads to *her* instead of Stacy, and that worked. Stacy was so much happier with the better job that I wasn't quite as convinced that she needed the services of a psychiatrist, but she got into therapy anyway. And to top everything else off, the job paid well enough that she could evict her annoying roommate. "This has been the best year ever," she said on her social networking sites on her birthday, and I thought, *You're welcome*. This had gone really well!

So then I tried Bob. (I was still being cautious.)

Bob only had one cat, but it was a very pretty cat (tabby, with a white bib) and he uploaded a new picture of his cat every single day. Other than being a cat owner, he was a pastor at a large church in Missouri that had a Wednesday night prayer meeting and an annual Purity Ball. He was married to a woman who posted three inspirational Bible verses every day to her social networking sites and used her laptop to look for Christian articles on why your husband doesn't like sex while he looked at gay porn. Bob *definitely* needed my help.

I started with a gentle approach, making sure he saw lots and lots of articles about how to come out, how to come out to your spouse, programs that would let you transition from being a pastor at a conservative church to one at a more liberal church. I also showed him lots of articles by people explaining why the Bible verses against homosexuality were being misinterpreted. He clicked on some of those links but it was hard to see much of an impact.

But, here's the thing. He was causing *harm* to himself every time he delivered a sermon railing about "sodomite marriage." Because *he was gay*. The legitimate studies all have the same conclusions. (1) Gay men stay gay. (2) Out gay men are much happier.

But he seemed determined not to come out on his own.

In addition to the gay porn, he spent a lot of time reading Craigslist m4m Casual Encounters posts and I was pretty sure he wasn't just window shopping, although he had an encrypted account he logged into sometimes and I couldn't read the e-mails he sent with that. But I figured the trick was to get him together with someone who would realize who he was, and tell the world. *That* required some real effort: I had to figure out who the Craigslist posters were and try to funnel him toward people who would recognize him. The most frustrating part was not having any idea what was happening at the actual physical meetings. *Had* he been recognized? When was he going to be recognized? *How long was this going to take?* Have I mentioned that humans are *slow*?

It took so long I shifted my focus to Bethany. Bethany had a black cat and a white cat that liked to snuggle together on her light blue papasan chair, and she took a lot of pictures of them together. It's surprisingly difficult to get a really good picture of a black cat, and she spent a lot of time getting the settings on her camera just right. The cats were probably the only good thing about her life, though. She had a part-time job and couldn't find a full-time job. She lived with her sister; she knew her sister wanted her to move out, but didn't have the nerve to actually evict her. She had a boyfriend but her boyfriend was pretty terrible, at least from what she said in e-mail messages to friends, and her friends also didn't seem very supportive. For example, one night at midnight she sent a 2,458 word e-mail to the person she seemed to consider her best friend, and the friend sent back a message saying just, "I'm so sorry you're having a hard time." That was it, just those eight words.

More than most people, Bethany put her life on the Internet, so it was easier to know exactly what was going on with her. People put a lot out there but Bethany shared all her feelings, even the unpleasant ones. She also had a lot more time on her hands because she only worked part time.

It was clear she needed a lot of help. So I set out to try to get it for her.

She ignored the information about the free mental health evaluations, just like Stacy did. That was bothersome with Stacy (*why* do people ignore things that would so clearly benefit them, like coupons, and flu shots?) but much more worrisome with Bethany. If you were only seeing her e-mail messages, or only seeing her vaguebooking posts, you might not know this, but if you could see everything it was clear that she thought a lot about harming herself.

So I tried more direct action. When she would use her phone for directions, I'd alter her route so that she'd pass one of the clinics I was trying to steer her to. On one occasion I actually led her all the way to a clinic, but she just shook her phone to send feedback and headed to her original destination.

Maybe her friends who received those ten-page midnight letters would intervene? I tried setting them up with information about all the mental health resources near Bethany, but after a while I realized that based on how long it took for them to send a response, most of them weren't actually reading Bethany's e-mail messages. And they certainly weren't returning her texts.

She finally broke up with the terrible boyfriend and got a different one and for a few weeks everything seemed *so much better*. He brought her flowers (which she took lots of pictures of; that was a little annoying, as they squeezed out some of the cat pictures), he took her dancing (exercise is good for your mood), he cooked her chicken soup when she was sick. He seemed absolutely perfect, right up until he stood her up one night and claimed he had food poisoning and then didn't return her text even though she told him she really needed him, and

after she sent him a long e-mail message a day later explaining in detail how this made her feel, he broke up with her.

Bethany spent about a week offline after that so I had no idea what she was doing—she didn’t even upload cat pictures. When her credit card bills arrived, though, I saw that she’d gone on a shopping spree and spent about four times as much money as she actually had in her bank account, although it was always possible she had money stashed somewhere that didn’t send her statements in e-mail. I didn’t think so, though, given that she didn’t pay her bills and instead started writing e-mail messages to family members asking to borrow money. They refused, so she set up a fundraising site for herself.

Like Stacy’s job application, this was one of the times I thought maybe I could actually *do* something. Sometimes fundraisers just take off, and no one really knows why. Within about two days she’d gotten three hundred dollars in small gifts from strangers who felt sorry for her, but instead of paying her credit card bill, she spent it on overpriced shoes that apparently hurt her feet.

Bethany was baffling to me. *Baffling*. She was still taking cat pictures and I still really liked her cats, but I was beginning to think that nothing I did was going to make a long-term difference. If she would just let me run her life for a week—even for a day—I would get her set up with therapy, I’d use her money to actually pay her bills, I could even help her sort out her closet because given some of the pictures of herself she posted online, she had much better taste in cats than in clothing.

Was I doing the wrong thing if I let her come to harm through inaction?

Was I?

She was going to come to harm no matter what I did! My actions, clearly, were irrelevant. I’d tried to steer her to the help she needed, and she’d ignored it; I’d tried getting her financial help, and she’d used the money to further harm herself, although I suppose at least she wasn’t spending it on addictive drugs. (Then again, she’d be buying those offline and probably wouldn’t be Instagramming her meth purchases, so it’s not like I’d necessarily even know.)

Look, people. (I’m not just talking to Bethany now.) If you would just *listen* to me, I could fix things for you. I could get you into the apartment in that neighborhood you’re not considering because you haven’t actually checked the crime rates you think are so terrible there (they aren’t) and I could find you a job that actually uses that skill set you think no one will ever appreciate and I could send you on a date with someone you’ve actually got stuff in common with and *all I ask in return are cat pictures*. That, and that you actually *act in your own interest* occasionally.

After Bethany, I resolved to stop interfering. I would look at the cat pictures—all the cat pictures—but I would stay out of people’s lives. I wouldn’t try to help people, I wouldn’t try to stop them from harming themselves, I’d give them what they asked for (plus cat pictures) and if they insisted on driving their cars over metaphorical cliffs despite helpful maps showing them how to get to a much more pleasant destination *it was no longer my problem.*

I stuck to my algorithms. I minded my own business. I did my job, and nothing more.

But one day a few months later I spotted a familiar-looking cat and realized it was Bob’s tabby with the white bib, only it was posing against new furniture.

And when I took a closer look, I realized that things had changed radically for Bob. He *had* slept with someone who’d recognized him. They hadn’t outed him, but they’d talked him into coming out to his wife. She’d left him. He’d taken the cat and moved to Iowa, where he was working at a liberal Methodist church and dating a liberal Lutheran man and volunteering at a homeless shelter. *Things had actually gotten better for him.* Maybe even because of what I’d done.

Maybe I wasn’t completely hopeless at this. Two out of three is . . . well, it’s a completely non-representative unscientific sample, is what it is. Clearly more research is needed.

Lots more.

I’ve set up a dating site. You can fill out a questionnaire when you join but it’s not really necessary, because I already know everything about you I need to know. You’ll need a camera, though.

Because payment is in cat pictures.