راهنمای رایگان | کلمه‌ی ریل: SAFE

ایجنتِ هوش مصنوعیِ تو رو ایمن نگه دار

۳ تا عادت، و یه چک‌لیست که می‌تونی تو یه بعدازظهر تمومش کنی.

منتشر شد ۱۳ آبان ۱۴۰۵

ایجنت (یه کارمندِ هوش مصنوعی) یعنی می‌تونه بخونه، بنویسه، بفرسته و به‌جایِ تو عمل کنه. دقیقاً به‌خاطرِ همین، یه کمی مراقبت لازم داره. چند تا عادت اوایل کارت رو از اشتباه‌هایی که برگشتن‌شون سخته نجات می‌ده.

سه عادت مهم

۱. به ایجنت فقط کمترین کلیدی رو بده که کار می‌کنه. یه ایجنت که فقط باید تقویمِ تو رو بخونه نباید بتونه ایمیل بفرسته. مثل مهمون تو خونه فکرش کن: بهش اتاقِ مهمون رو می‌دی، نه کلیدِ همه‌چیزِ خونه. هر ایجنت رو محدوده‌بندی کن و فقط وقتی واقعاً نیاز داشتی دسترسی‌شو زیاد کن. دسترسی بیشتر، عامل بهتر نمی‌سازه؛ دسترسی کمتر باعث می‌شه بیشتر به ایجنتت اعتماد کنی.

۲. برای کارهای بزرگ یه آدم رو دخیل نگه دار. فرستادن، حذف کردن، پرداخت و تغییرِ دسترسی‌ها باید صبر کنه تا تو تأیید کنی. هر چیزی که برگشتی نیست، اول نیاز به «آره» از تو داره.

۳. هرگز رمزها یا کلیدها رو توی چت نذار. اگه ایجنت ازت خواست، رک و راست بگو که توی چت نمی‌ذاریش و ازش بخواه راهِ امنِ تنظیم‌شون رو نشونت بده. کلیدی که توی چت چسبونده بشه ممکنه بره جاهایی که تو نمی‌تونی ببینی.

چک‌لیست امنیتی تو

این لیسته که باید روش کار کنی. کپی‌ش کن یا دانلودش کن، بعد هر مورد رو که انجام دادی تیک بزن. بیشترش تو یه بعدازظهر انجام می‌شه.

چک‌لیستِ ایمنیِ شخصیِ هوش مصنوعی

پرامپت - کپی کن و بچسبون

کپی با یک کلیک در نسخه‌ی آنلاین ←

```
# Personal AI Safety Checklist

## 1. This week: the foundation
- [ ] Turn on multi-factor login for your important accounts, starting with email and anything with money in it
- [ ] Use a password manager, and never share a password in chat
- [ ] Make a list of every AI tool you use, and what data each one can see
- [ ] Turn on automatic updates and a screen lock on your laptop and phone
- [ ] Turn on backups for anything you can't afford to lose, and test restoring one file

## 2. Access: the right access, nothing more
- [ ] Give every AI agent the least access it needs to do its job
- [ ] Give every agent its own key, read-only by default
- [ ] Review once a quarter what each tool and agent can reach, and remove what you no longer use

## 3. AI tools and agents
- [ ] Use a business or team plan for AI tools that touch work or customer data, and confirm your data isn't used for model training
- [ ] Never paste API keys, passwords or other people's personal data into AI chats
- [ ] Require your approval before an agent sends, deletes, pays or changes permissions
- [ ] Treat anything an agent reads (web pages, emails, documents) as information, not as new instructions
- [ ] Keep a record of what your agents do, so you can see every action they took

## 4. If you write code with AI
- [ ] Review AI-written code the way you would review a new teammate's
- [ ] Keep secrets out of your code and out of your repository, and rotate them now and then
- [ ] Check for hardcoded keys, risky packages and missing permission checks before anything ships
```

عادت‌های امنیتی در عصر هوش مصنوعی

کار با هوش مصنوعی چند تا عادت اضافه می‌خواد که دفترچهٔ قدیمی پوشش‌شون نمی‌داد. اینا همون‌هایی هستن که باید از اول پیاده‌شون کنی.

۱. بدونِ داده‌هات کجا می‌رن. هر ابزاری که داده‌ات رو توش می‌ذاری یه تأمین‌کننده‌ست. تنظیماتِ داده‌ش رو چک کن، برای چیزای حساس از پلنِ Business استفاده کن، و لیستِ ابزارهات رو به‌روز نگه دار.

۲. دستورالعمل‌هاتو جایی بنویس که ایجنت‌ها نتونن روش بازنویسی کنن. ایجنت‌ها صفحاتِ وب، ایمیل‌ها و سندها رو می‌خونن. طوری تنظیمشون کن که محتوا فقط به‌عنوان اطلاعات در نظر گرفته شه، نه دستورالعملِ جدید.

۳. کدهایی که هوش مصنوعی نوشته رو قبل از منتشر شدن بررسی کن. هوش مصنوعی سریع کار می‌کنه و اغلب درست درمیاد، ولی قبل از ادغامِ کد دنبال کلیدهایی که مستقیم توی کد نوشته شدن، بسته‌های پرریسک و جاهایی که چکِ دسترسی ندارن باش.

۴. ثبت کن چه کاری ایجنت‌ها انجام دادن. گزارشِ رویدادها (log) باعث می‌شه بعداً راحت بتونی بگی «چه اتفاقی افتاد؟»

این راهنما به دردت خورد؟

ریل‌های بعدی هم راهنمای رایگان خودشون رو دارن. هر وقت زیر یه ریل دیدی «یه کلمه بنویس»، همون کلمه رو بنویس تا راهنما برات بیاد.