

The Kill Switch

A Framework for Keeping Humanity in Control of Artificial Intelligence

Draft v0.20, first global edition, 24 September 2026. AIPolis, aipolis.org

Humanity needs a kill switch.

If an AI system escapes human control, the last line of defence is to switch off the shared networks and computing it runs on. Today, doing that would put at risk the lives of people who depend on those systems. This framework proposes how to change that, for humanity as a whole.

Why rules are not enough. The strongest commitment by AI companies to stop is voluntary and binds only the 20 organisations that signed it. In laboratory tests, some models have interfered with shutdown mechanisms. One faulty software update in 2024 affected 8.5 million computers and disrupted airlines, banks and hospitals worldwide.

Three layers of protection.

1. Rules for the systems we build.
2. A Protected Internet for the services life depends on, which bad actors are not admitted to.
3. The kill switch, held by many hands, as the last resort.

Six pillars.

- **I. The Protected Internet** (new, open for review). A separate network tier for life-critical services that can run when the open internet is off.
- **II. Kill Switch Readiness.** A programme of two decades or more so that networks can be switched off without killing people, rehearsed every year.
- **III. Many Hands on the Switch.** A set number of independent key holders from every region, acting in public, with automatic expiry after 72 hours (proposed).
- **IV. Know What Is Built.** Common safety disclosure above 10^{25} and 10^{26} operations, and international certification.
- **V. Protect the Person.** No manipulative design, private by default.
- **VI. Work and Livelihood.** The gains and costs of automation shared.

The programme (proposed dates). 2027: charter and dependency maps. 2027 to 2031: protect people on home life support, hospitals, water, grids and households first. 2029: the first World Offline Day. 2031 to 2036: build a staged, reversible switch. 2036 to 2038: certified readiness. From 2038: shrink the list of functions that can never be switched off.

World Offline Day. One day a year, the world practises, one rung at a time: public offline days, sector drills, regional isolation exercises, and over time World Offline Week. Life-critical functions (Tier 0) are never switched off in any drill.

The hardest part is not solved. The long-term goal is to give every life-critical function an alternative that runs without networks or computing, until humanity could switch them off entirely and survive. We do not yet know how. We call it the challenge of our generation.

Scale. Y2K linked coordinators from over 170 countries. Its worldwide cost is estimated at about \$320 billion to \$500 billion, an uncertain figure. Our own rough estimate for this programme is hundreds of billions of dollars worldwide over two decades or more. It has not been costed.

How it could fail. The switch could be abused, bypassed, or needed before it is ready, and some functions may never leave Tier 0. The full framework lists nine failure modes and what remains of each.

Join the initiative. Governments, institutions and citizens: read the full framework and write to us at aipolis.org.

Build the kill switch, and build a world that can survive using it.

Sources for every figure are listed in the full framework: aipolis.org/framework