NOVARA
CONSULTING GROUP

This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/04/18/when-control-systems-inherit-the-same-uncertainty-they-are-meant-to-govern/.

Thaum Cov Kev Tswj Xyuas (Control Systems) Yig Tau Tib Yam Kev Tsis Ntseeg Siab Uas Lawv Yog Los Tswj

Heather Grizzle · April 18, 2026

When AI guardrails are themselves probabilistic models, governance inherits the uncertainty it is meant to control. What real control requires.

Tam sim no lub tswv yim tswj xyuas AI (AI governance) tab tom raug piav thiab muag raws li yam yooj yim tshaj kom pom tau txoj kev xav tsis yog raws li yam tseem ceeb tshaj kom soj ntsuam.

Kev sib tham no tau raug kav los ntawm cov khoom pom tau tseeb: cov dashboard tswjfwm, cov qauv tswj sib txheeb, cov guardrail muaj npe, thiab cov txheej txheem ua raws theem uas qhia tias ib qho kev thov tau raug soj ntsuam kom meej ua ntej yuav muaj kev nqis tes ua. Cov qauv sawv cev no tsim kev xav tias muaj kev tswj hwm vim lawv zoo sib xws li cov kev saib xyuas system qub. Txawm li cas los, lawv npog lub teeb meem tseem ceeb dua ntawm hom kev ua haujlwm twg tiag tiag tab tom ua qhov kev ntsuam xyuas nyob rau hauv cov txheej ntawd.

Nyob rau hauv feem ntau cov qauv tsim tam sim no, ib lub AI system tsim ib qho kev nqis tes ua raws li kev thov, thiab qhov kev thov ntawd raug xa mus rau ib txheej txheem ntsuam xyuas uas thaum kawg yog lwm lub qauv txiav txim seb qhov kev nqis tes ua ntawd puas yuav tsum txuas ntxiv. Txheej kev ntsuam xyuas no tuaj yeem raug piav qhia txawv txawv nyob ntawm tus muag khoom lossis lub qauv siv. Nws tuaj yeem raug muab tso rau hauv qhov chaw ua ib lub cav tswj cai, ib lub cuab yeej ntsuas kev pheej hmoo, lossis ib lub kaw lus ntsuas kev ntseeg siab uas raug txhawb nqa nrog cov cai. Txawm hais tias muaj qhov txawv ntawm cov lus siv, tab sis lub hauv paus qauv tseem yog ib yam khiav khiav. Ib lub qauv sib npaug (probabilistic model) raug muab txoj haujlwm ntsuam xyuas seb qhov tsim tawm los ntawm lwm lub qauv sib npaug puas tsim nyog.

Qhov kev xaiv tsim khoom no muaj lub ntsiab lus loj thaum soj ntsuam los ntawm qhov saib pheej hmoo thiab kev ntseeg siab tau. Cov qauv sib npaug, raws li lub ntsiab lus, tsis tsim cov txiaj ntsig deterministic, rov ua tau ib yam nyob rau hauv tib yam xwm txheej. Cov txiaj ntsig lawv tsim tawm raug tswj los ntawm cov qauv statistical uas tau muab los ntawm cov ntaub ntawv qhia, thiab cov kev ntsuam xyuas lawv kuj raug cuam tshuam ib yam los ntawm kev xav sib npaug tsis yog los ntawm cov cai ruaj khov. Thaum ib lub system zoo li no raug siv los tswj lwm lub system uas yog hom tib yam, txoj haujlwm tswj hwm ntawd tsis tau qhia txog ib hom kev ntseeg siab tau tshiab. Es tsis txhob li ntawd, nws rov ua tib qho kev tsis ntseeg tau ntawd dua nyob rau ib theem siab dua ntawm kev nkag siab.

Los ntawm qhov saib tswj hwm, qhov no tsim ib lub qauv sib vov. Lub system uas yog npaj los tswj cai tab tom raug ceev faj rau tib yam kev hloov, kev tsis ruaj khov ntawm cov xwm txheej txawv txawv, thiab kev nyaj rau ntsiab lus ib yam li lub system uas nws tab tom saib xyuas. Yog li ntawd, qhov muaj cov txheej txheem ntxiv, kev kuaj xyuas, lossis cov khoom qhwv ntxiv tsis txhais tau tias muaj kev txo pheej hmoo raws lub ntsiab. Nws tuaj yeem ua rau muaj kev nyuaj tshaj qub, thiab nws tuaj yeem tsim tau qhov zoo li muaj kev ceev faj, tab sis nws tsis hloov tau tus cwj pwm hauv paus ntawm lub system kom tsim tau ib qhov ciam teb tswj cai tshwj xeeb.

> Nws qhia txog kev rov qab (recursion) tsis yog kev tswj hwm (control).

Qhov txawv no yuav pom meej dua thaum muab piv rau seb kev tswj hwm ua haujlwm licas nyob rau hauv lwm cov chaw pheej hmoo siab. Piv txwv li, hauv cov qauv nyiaj txiag, cov txheej txheem pom zoo kev sib pauv nyiaj txiag muaj kev tso siab rau cov cai twb muaj los ua ntej, cov ntaub ntawv audit, thiab cov qauv tib neeg sau uas tuaj yeem taug qab thiab rov txhim khu tau. Nyob rau hauv cov qauv engineering uas pheej hmoo siab thiab tseem ceeb, cov qauv tswj hwm raug tsim los ua haujlwm nrog cov khoom ntseeg siab tau txawv ntawm cov qauv uas lawv saib xyuas. Lub cuab yeej tswj hwm raug faib tawm rau ntawm ob qho tib si txoj haujlwm thiab tus cwj pwm kom nws tuaj yeem muab ib qhov chaw ruaj khov rau kev xa mus xam thaum muaj kev tsis ntseeg tau lossis kev ua tsis tiav tshwm sim.

Cov qauv tswj hwm AI tam sim no feem ntau tsis tau tsim qhov kev cais tawm ntawd. Es tsis txhob li ntawd, lawv muab txiaj ntsig kev txiav txim sib npaug tso rau saum kev tsim tawm sib npaug thiab suav qhov qauv tshwm sim ua ib lub system tswj hwm.

Txoj kev no xav tias yog ntxiv cov theem ntsuam xyuas ntxiv nyob rau hauv tib chav kev ntseeg siab tau yuav tsim tau cov txiaj ntsig kev tswj hwm. Raws li tiag tiag, nws qhia txog kev rov qab tsis yog kev tswj hwm, vim tias txhua txheej nyob ntawm tib hom kev xav thiab yooj yim raug cuam tshuam los ntawm tib hom kev ua tsis tiav.

Qhov tshwm sim tsis yog xav xwb. Thaum ib lub system tsim ib qhov txiaj ntsig ua rau muaj kev puas tsuaj, lo lus nug uas nqis tom qab tsis yog seb puas muaj ntau txheej nyob rau hauv txoj kev txiav txim. Lo lus nug yog seb puas muaj ib txheej twg ua haujlwm ua ib qhov tswj hwm tseeb, txhais tau tias nws ua haujlwm nrog ib theem deterministic, kev taug qab tau, thiab kev lav ris uas txawv ntawm lub system nws tab tom tswj hwm. Yog txhua txheej muaj tib lub hauv paus sib npaug, ces lub system tsis muaj ib qhov chaw meej uas kev lav ris tuaj yeem raug tso ruaj lossis kev txiav txim tuaj yeem raug rov txhim khu tau ntseeg siab tau.

> **QHOV KEV SIM QHIA TSEEB**
>
> Kev tswj hwm raug txhais los ntawm kev tsim ib lub cuab yeej ua haujlwm nrog ib theem kev ntseeg siab tau txawv ntawm lub system uas raug tswj hwm.

Txoj kev pom uas nqis tawm los ntawm txoj kev soj ntsuam no meej tshwm sim ncaj qha. Kev tswj hwm tsis raug txhais los ntawm tus lej ntawm cov kev tswj cai muaj nyob rau hauv ib lub system, thiab tsis raug txhais los ntawm kev pom tau ntawm nws cov qauv tsim khoom. Kev tswj hwm raug txhais los ntawm kev tsim ib lub cuab yeej ua haujlwm nrog ib theem kev ntseeg siab tau txawv ntawm lub system uas raug tswj hwm. Yog tsis muaj qhov txawv ntawd, cov txheej ntxiv tsis yog kev tswj hwm. Lawv tsuas yog tsav tib yam kev tsis ntseeg tau dhau ntau lub feem ntxiv.

Thaum ib txheej tswj hwm ua haujlwm zoo ib yam li lub system nws yog npaj los tswj cai, cov qauv tsim khoom tsis tau ua tiav kev tswj hwm raws lub ntsiab lus tseeb. Nws tsim ib lub qauv rov qab uas kev tsis ntseeg tau ntsuam xyuas kev tsis ntseeg tau, thiab ciam teb ntawm kev nqis tes ua thiab kev saib xyuas puas tsuaj. Qhov puas tsuaj ntawd yog qhov chaws pheej hmoo tiag tiag nyob, tsis hais lub system saib zoo li muaj qauv npaj tseg npaum li cas rau sab nraud.

TIPS

Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

Yuav Hais Qhov Chaw Li Cas

Grizzle, H. M. (2026, April 18). Thaum Cov Kev Tswj Xyuas (Control Systems) Yig Tau Tib Yam Kev Tsis Ntseeg Siab Uas Lawv Yog Los Tswj. Novara Consulting Group. https://www.novaracg.com/hmn/2026/04/18/when-control-systems-inherit-the-same-uncertainty-they-are-meant-to-govern/