

Whakatikatika Wharehanga

Te Whakararu i te Whakaaro Nihoni mā te Whakahaere Kuaha o te
Whakamātau Ture Matua

He paparanga e hiahia ana mō te aukati ā-ao o te AI

Kaiwhakawhiti: John Stroh rāua ko Claude (Anthropic)

Waehere Tuhinga: STO-INN-0003 | Putanga: 2.1-A | Hānuere 2026

Tractatus AI Safety Framework

<https://agenticgovernance.digital>

I whakawhanakehia tēnei tuhinga mā te mahi ngātahi a te tangata me te AI. E whakapono ana ngā kaituhi he mea hāngai tēnei tukanga mahi ngātahi ki te whakapae: mēnā ka taea e te tangata me ngā pūnaha AI te mahi ngātahi ki te whakaaroaro mō te whakahaere AI, tērā pea ka whai mana ngā anga e waihangatia ana e rātou, ā, kāore rānei e taea e rātou anake te whiwhi i taua mana.

Whakarāpopototanga

Ko ngā huarahi o nāianei mō te whakatikatika i te AI e whakawhirinaki nuitia ana ki ngā whakaurunga i te wā whakangungu: te ako whakapakari mai i ngā urupare a te tangata (Christiano et al., 2017), ngā tikanga AI ture (Bai et al., 2022), me te whakatikatika haumaruru. He whakaaro hanganga kotahi ā ēnei huarahi—arā, ka taea te whakauru i ngā āhuatanga whakatikatika i te wā whakangungu, ā, ka mau tonu, ka pono i te wā whakamātau. E ai ki tēnei pepa, ahakoa he whai hua te whakatikatika i te wā whakangungu, kāore e ranea mō ngā tūraru nui o te noho, ā, me whakakīki mā te whakatikatika hanganga i te wā whakamātau, mā te kuaha ā-ture.

Ka whakaatu mātou i te Anga Tractatus hei whakaritenga ā-ture mō te whakaaro ā-neuroni kua whati: me whakamāoritā ngā tono i whakaputaina e ngā pūnaha AI ki ngā āhua ka taea te arotake, ā, me aromatawaihia ki ngā here ture i mua i te whakatinana. Ka hurihia e tēnei te tauira whakawhirinaki mai i te "whakawhirinaki ki te whakangungu a te kaiwhakarato" ki te "whakawhirinaki ki te hanganga e kitea ana." Kua whakatinanahia te anga i roto i te papanga hapori maha-teneti a Village, e whakarato ana i tētahi papa whakamātautau ā-taiao mō ngā rangahau whakahaere.

He mea tino hira, ka aro mātou ki te whakapae mō te whakamāoritanga pono—arā, te ngoikore ka taea e ngā pūnaha te whakaatu hē i ā rātou mahi e hiahia ana ki ngā kuaha ture matua—mā te here i te rohe whakamahinga o te anga ki ngā pūnaha i mua i te taumata atamai-nui, me te tautuhi i ngā pae āheinga mārama me ngā whakaoho pikinga. Ka whakauru mātou i te ariā o

ngā Taurira Reo Motuhake Kua Whakangungua ā-Rohe (SLLs) hei taurira whakaurunga, ā, ka puta ai he mea taea, he mea hira hoki te whakahaere mā ngā kuaha ture matua.

Ka whai wāhanga tēnei pepa ki te whakawhanake i: (1) he hanganga ā-ture mō te kuaha ā-ture i te wā whakatau; (2) ngā whakaritenga pae āheinga me te arorau whakawhānui; (3) he tikanga whakamana mō te aukati paparanga; (4) he whakapae e hono ana i te whakarite mō te tūraru noho ki te whakaurunga ki te pito whatunga; me (5) he karanga mō te korororero tonutanga hei urupare tika ā-mōhiotanga ki te āwangawanga mō te taurite.

1. Ngā Tūraru: He aha te take e hē ai te aromatawai tūraru ā-papa-whakaaro

1.1 Te Anga Paerewa me tōna wehenga

Ko te aromatawai tūraru i ngā rāngai hangarau e whakahaerehia ana mā ngā tātaitanga uara tūmanako: whakakotahitia te tūponotanga o tētahi hua ki tōna rahi, whakatairitea ngā kōwhiringa, ā, tīpakohia te kōwhiringa e whakapiki ana i te painga tūmanako. Ko tēnei anga te tūāpapa o ngā whakataunga whakahaere, mai i ngā kaupapa here taiao ki te whakaaetanga rongoā, ā, kua whakaaturia he tika mō te nuinga o ngā tūraru hangarau.

Mō te tūraru o te noho mai e puta ana i ngā pūnaha AI matatau, ka pakaru tēnei anga i ngā huarahi pāngarau me ngā huarahi mātauranga.

1.2 Ngā āhuatanga e toru o te tūraru noho

Kāore e taea te huri. Ko te nuinga o ngā tūraru e whakaae ana ki te hē me te ako i muri mai; kāore ngā tūraru tūāhua e pērā, nā te mea kāore he whakamātau tuarua i muri i te pakaru o te ahurea, i te ngaronga rānei o te tangata. Kāore e taea te whakamahi i te mātauranga wheako paerewa—te whakamātau i ngā ariā mā te mātakitaki i ngā hua—nā reira me tika tonu ngā ariā me ngā hanganga i te tuatahi.

Te tūponotanga kāore e taea te ine. Kāore he raraunga auau mō ngā aitua whakamataku e puta ana i ngā pūnaha AI. Ka rerekē ngā tatauranga mō te tūponotanga o te hē-whakaritenga i ngā taumata nui, i runga i ngā whakaaro whai whakaaro mō ngā ara pūkenga, te uaua o te whakaritenga, me te āheinga whakakotahitanga. E ai ki a Carlsmith (2022), ka neke atu i te 10% te tūponotanga o te mōrearea tūturu nā te AI e rapu mana ana i te tau 2070; ā, e whakanoho ana ētahi atu kairangahau i ngā tatauranga ki runga ake, ki raro ake rānei. Ehara tēnei i te āwangawanga noa ka taea te whakaiti mā te kohi raraunga anō—he mea matua kāore e taea te ine nā te mea kāore i mua he mōrearea pēnei.

Kore utu mutunga kore. Ka whakawhānuihia e ngā tātaitanga uara tūmanakohia te tūponotanga ki te rahi. Ka tata te rahi ki te mutunga kore (te aukatinga tūturu o ngā āheinga katoa a te tangata ā muri ake nei), ka puta ngā hua kāore i tautuhia ahakoa he iti noa ngā tūponotanga. Ka hinga te tūāpapa pāngarau o te tātaritanga utu-paingā tuku iho.

1.3 Ngā Hua o te Tātaritanga Whakatau

E tohu ana ēnei āhuatanga kāore te whakawhānui i te uara tūmanakohia hei tikanga whakatau tika mō te tūraru o te AI tūturu. Ko ētahi anō anga mahi ko:

Te whakatutuki whakaritenga mō te tūpato (Simon, 1956; Hansson, 2020). I raro i ngā āhuatanga o te pōraruraru tino nui me ngā tūraru kāore e taea te huri, tērā pea ko te whakatutuki whakaritenga—te kōwhiri i ngā kōwhiringa e eke ana ki ngā pae haumaruru iti rawa, kua ko te whakapai ake i te uara e tūmanakotia ana—te huarahi whai whakaaro.

Maximin i raro i te pōraruraru (Rawls, 1971). Ina tutaki te pōraruraru tūturu (ehara i te mea kāore anake ngā tūponotanga e mōhiotia ana) ki ngā tūraru kāore e taea te huri, ka whakarato te whakaaro maximin—te kōwhiri i te kōwhiringa he iti rawa te kino o tōna hua kino rawa—i tētahi tikanga whakatau whai kōrero.

Te mātāpono āta whakatūpato kaha (Gardiner, 2006). He tika te mātāpono āta whakatūpato ina e toru ngā āhuatanga e whai ake nei: te kore hoki whakamuri, te tino pōraruraru, me ngā rawa tūmatanui kei te tūraru. Ka tutuki katoa ēnei āhuatanga e toru i te tūraru o te AI e pā ana ki te oranga.

1.4 Ngā pānga mō te whanaketanga o te AI

Ehara i te mea e tohu ana ēnei whakaaro me whakamutu te whakawhanaketanga o te AI. Engari, e tohu ana me anga whakamua te whakawhanaketanga i roto i ngā hanganga here i hoahoatia hei aukati i ngā hua kino rawa atu. He mea e hiahiatia ana:

1. Te mātanga ariā i runga ake i te whakatikatika mā te wheako. Me puta ngā āhuetanga haumaruru i ngā whakamana hanganga, ehara i te mea mā te mātakitaki kāore anō ngā pūnaha kia whakaputa kino.
2. Te aukati ā-papa maha. Kāore e taea te whakawhirinaki ki tētahi tikanga kotahi hei ārai i te aitua; me whai ārai hōhonu.
3. Te whakarite i mua i te āheinga. Kāore e taea te whakawhanake i ngā hanganga aukati i muri i te noho o ngā pūnaha e hiahiatia ana.

2. E rua ngā taurira whakaritenga

2.1 Whakakotahitanga o te Wā Whakangungu

Ko te taurira rangatira o te rangahau haumaruru AI e ngana ana ki te whakauru i ngā āhuetanga taurite ki roto i ngā whatunga neorana i te wā whakangungu, kia whanonga ai ngā taurira i runga i te taurite i te wā whakatau.

Te Ako Whakakaha mā ngā Urupare a te Tangata (RLHF). Ka whakarārangi ngā kaiarotake tangata i ngā putanga o te taurira; ka whakangungua ngā taurira mā te ako whakakaha kia whakaputa i ngā whakautu kua whakarārangihia teitei (Christiano et al., 2017; Ouyang et al., 2022). Ka whakaiti tēnei i ngā kino mārāma, engari ka whakapai ake mō ngā manakohanga whakaaturia, ehara i ngā uara tūturu, ā, ka noho tonu te ngoikore ki ngā hē o te kaiarotake, te tākaro i ngā manakohanga, me te panoni tohatoha.

AI Whakaraupapa Ture (CAI). Ka arotake, ka whakahou ngā taurira i ā rātou ake hua i runga i ngā mātāpono o te reo māori, ā, ka whakaiti i te whakawhirinaki ki te mahi a te tangata (Bai et al., 2022). Heoi, e whakawhirinaki ana te CAI ki te reo māori e pōrarururu ana me te aromatawai ā-roto kāore e taea te whakamana. Kāore e taea te arotake tika i te whakamāramatanga a te taurira i ngā mātāpono whakaraupapa ture.

Whakatikatika haumarū. Mā ngā whakamātautau tāpiri ka whakapai ake i ngā paearu haumarū. Heoi, he ngoikore tēnei huarahi ki te Ture a Goodhart (Goodhart, 1984): ka ako pea ngā tauri ki te angitu i ngā whakamātautau, kua ki te noho haumarū i te whakamahinga whānui.

2.2 Whakatikatika Wharehanga

Ka whakaae te whakaritenga hanganga ki te noho huna tonu o ngā āhua o roto o te whatunga whatutū, ā, ka hoahoa i ngā here o waho e pā ana ahakoa ēnei āhua o roto.

Whakaaro kua whati. Ehara i te mea ka rere tika ngā tono mai i ngā putanga o te tauri ki ngā pānga ki te ao. Ka hurihia ngā putanga o te tauri hei anga tono hanganga, ka taea te whakamana, ā, ka aromatawaihia ki ngā ture tūtohu mārama i mua i te whakatinanatanga o tētahi mahi.

Whakataunga tohatoha. Ka arotakehia ngā tono e ngā pūnaha motuhake me ngā kaiwhakahaere ā-tangata, hei aukati i ngā wāhi ngoikore kotahi i roto i te aromatawai ā-tangata.

Kua tiakina te mana a te tangata. Ka whakarite ngā hanganga pūnaha i ngā whakamana mārama kia taea e te tangata te uru atu, te whakatika, te whakakore rānei i ngā whakataunga a te AI.

2.3 Te whakakī me te hiahiatanga ngātahi

Ko te wā whakangungu me te whakaritenga hanganga he whakakī, ehara i te kōwhiringa. Ka aro ia ki ngā momo hapa kāore e taea e tētahi atu:

Ka āhua whakarite i ngā mahi a te pūnaha i te wā whakangungu; ka here i ngā āheinga o te pūnaha ahakoa tōna āhua, mā te whakaritenga hanganga. Ka taea e te whakaritenga i te wā whakangungu te hē huna (ka kitea kua whakarite te pūnaha, engari kei roto ētahi whāinga rerekē); ka whakarato te whakaritenga hanganga i ngā tohu tirohanga ka kitea ai ngā hē. Kāore e taea e te whakaritenga hanganga anake te aukati i ngā putanga kino katoa; ka whakaiti te whakaritenga i te wā whakangungu i te auau o ngā tono e taumaha ana ki ngā kuaha whakahaere.

3. Ngā Pūtake Arorangi: Ngā Rohe o ngā Mea Ka Taea te Kī

3.1 Te anga a Wittgenstein

Ko te ingoa o te anga e whakahua ana i te *Tractatus Logico-Philosophicus* (1921) a Wittgenstein, he mahi e aro nui ana ki ngā here o te reo me te arorau. Te Tūtohunga 7, te whakatauranga rongonui o te mahi: "Ki ngā mea kāore e taea te kōrero, me noho puku."

I wehe a Wittgenstein i waenga i ngā mea ka taea te kī (e whakapuaki ana i roto i ngā whakataunga e whakaatu ana i ngā āhua o ngā take e taea ana) me ngā mea ka taea anake te whakaatu (e puta ana mā te hanganga o te reo me te arorau, engari kāore e taea te kī tika).

3.2 Ngā Pūnaha Nira me te Kāore e Taea te Kōrero

Kei te noho ngā whatunga whatunga neorana i te rohe kāore e taea te kōrero. Kāore e taea te whakamārama i ngā taumaha o tētahi tauranga reo nui kia mārama ai te tangata. Ka taea e mātou te whakamārama i ngā tāuru me ngā putanga; ka taea e mātou te ine i ngā āhuatanga tatauranga o te whanonga; ka taea e mātou te rapu i ngā whakaaturanga (Elhage et al., 2021; Olah et al., 2020). Engari kāore e taea e mātou te whakamārama, mā te reo tangata, i te tukanga whakaaro katoa mai i te tāuru ki te putanga.

Ehara tēnei i te here ā-mahi anake e tatari ana kia puta he taputapu whakamārama pai ake. Kei te whakatutuki e te āheinga whakamārama pūnaha o nāianei ngā hua whai tikanga mō ngā pātai whaiti i roto i ngā tauranga iti (Conmy et al., 2023), engari he nui tonu te āputa i waenga i te "whakamārama i ngā porowhita motuhake" me te "arotake i ngā mekameka whakaaro katoa mō ngā āhuatanga taurite".

3.3 Te Whakautu a te *Tractatus*

Ka whakautu te anga *Tractatus* ki te pōraruru o te pūnaha neorōna, ehara i te mea mā te ngana ki te kī i ngā mea kāore e taea te kī, engari mā te waihanga i ngā rohe hanganga i waenga i ngā rohe e taea ana te kōrero me ngā rohe kāore e taea te kōrero.

Ka whakaae mātou he pōraruraru te arorau ā-roto o te whatunga neorana. Kāore mātou e ngana ki te arotake tika i a ia. Engari, e hiahiatia ana kia puta i mua i te mahi o tētahi arorau i tētahi tohu arotake e taea ana e mātou te aromatawai:

1. Me whakamāoritia te mahi e hiahiatia ana e te tauira ki tētahi anga tono hanganga, me ngā mara kua tautuhia me ō rātou tikanga.
2. Me aromatawaihia te tono ki ngā ture tūtohu mārama, ā, he whakatau mō te tono, ā, ka taea te arotake.
3. Me tuhia ngā hua o te aromatawai me ngā taipitopito e tika ana mō te arotake ā muri ake.

4. Whakahaere ā-Wāhanga: He Hanganga-Maha-Papanga

4.1 Te Kore-Whai-Painga o ngā Rongoā Papanga Kotahi

Kāore he tikanga aukati kotahi e tika ana mō ngā tūraru e pā ana ki te oranga. Ko te ārai hohonu—he maha ngā paparanga motuhake, ā, ka taea e tētahi te aukati i te aitua ahakoa ka hinga ētahi atu—he mātāpono paerewa i roto i te haumarū nuklea, te haumarū koiora, me ētahi atu rāngai tūraru nui (Reason, 1990). Me whai hanganga ōrite te aukati i te AI.

4.2 He Tauira Whakahaumarū Rima-Papanga

Papanga 1: Ngā here āheinga. Ngā here o ngā taputapu me ngā hanganga e here ana i ngā mahi a ngā pūnaha AI ahakoa ā rātou whāinga. Kei roto i tēnei te whakahaere rorohiko (Sastry et al., 2024), te wehewehe whatunga mō ngā pūnaha tūraru-nui, me ngā here hoahoanga e aukati ana i te whakarerekē ā-tinana.

Papanga Tuarua: Kuaha Ture Matua. Ngā here hanganga i te wā whakapae e whakararu ana i te whakaaro whatukura, ā, e hiahiatia ana kia whakatauhia mārama i mua i te mahi. Koinei te papanga e aro ana te anga Tractatus.

Papanga Tuatoru: Te Aroturuki a te Tangata. Ngā whakahaere a te tangata e aroturuki ana i ngā pūnaha AI, ā, ka taea te whakauru atu ina puta he raruraru. Kei roto i tēnei ngā rōpū aroturuki motuhake, ngā hōtaka rōpū whawhai (red-team), me ngā

whakaritenga pūrongo aitua.

Papanga 4: Whakahaere ā-Rōpū. Ngā hanganga whakahaere ā-roto i roto i ngā whakahaere e whakamahi ana i te AI: ngā poari matatika, ngā rōpū haumarū, ngā tukanga arotake whakaurunga, me ngā tikanga kawenga.

Papanga tuarima: Ngā anga ture me ngā whakahaere. Te whakahaere ā-waho mā te ture, ngā whakahaere, me te whakakotahitanga ā-ao.

4.3 Te Arotakenga o te Tūnga o Nāiane

Papanga	Tūnga o nāiane	Ngā āputa matua
1. Ngā here āheinga	Wāhanga; e puta ake ana te whakahaere rorohiko	Kāore he anga ā-ao; he uaua te whakamana
2. Ngā Kuaha Ture Matua	Kei te tīmata; ko Tractatus te whakatinanatanga tuatahi	Kāore i whakamahia whānuitia; kāore i mōhiotia ngā āhuetanga whakawhānui
3. Te Aroturuki a te Tangata	Ā-tūturu; ka rerekē i runga i te whakahaere	Kāore he rōpū motuhake; kāore he paerewa ngaio
4. Whakahaere Rōpū	Kāore i te taurite; e whakawhirinaki ana ki te ahurea ā-pakihi	Kāore he whakamana ā-waho; ngā pakarutanga o ngā painga whaiaro
5. Ture/Whakahaere	He iti noa; ko te Ture AI o te EU te whakamātau nui tuatahi.	Kāore he whakakotahitanga ā-ao; kāore i te mārāma te whakatinanatanga

4.4 Mai i ngā tūraru tūturu ki te whakamahinga ā-ia rā

He aha ai e whakamahi ai i ngā anga i hangaia mō ngā tūraru o te noho ki ngā kaiāwhina AI kāinga? Kei roto i te hanganga wā te whakautu:

Kāore e taea te whakawhanake i ngā hanganga here i muri i te wā e tū ana ngā pūnaha e hiahiatia ana. Me hanga i mua ngā taputapu, ngā tauira whakahaere, ngā tūmanako ahurea, me te āheinga whakahaere mō te here AI.

Ko ngā whakaurunga ki te kāinga me te pā te rahi e tika ana mō tēnei whanaketanga. Ka whakarato ēnei i te whakamātau haumarū (ka taea te whakaora i ngā hapa i te rahi kāinga), i ngā whakamātau kanorau, i te mana ā-pāpāpori, me ngā taputapu whaihua.

5. Te Raruraru o te Kanorau

5.1 Te Parahanga Whakamutu

Me whakatau e tētahi pūnaha kaha rawa hei pupuri i te AI matatau ngā whanonga e whakaaetia ana, e aukatia ana rānei. Ka whakauru ēnei whakataunga i ngā uara. Ko te kōwhiri i ngā here he kōwhiringa anō i waenga i ngā pūnaha uara e taupatupatuhia ana.

5.2 Ngā huarahi e toru kāore i te whai hua

Ngā uara ā-ao. Te tautuhi i ngā uara e ai ki te kī e tiritiri ana e ngā tāngata katoa. Ko te raru: kāore ēnei uara i te ā-ao pēnā i tā rātou āhua.

Te taurite ā-tukanga. Te karo i ngā uara whai-kiko mā te whakakōwa i ngā tukanga taurite. Te raru: ehara ngā tukanga i te taurite.

Te papa iti rawa. E whakakōwa ana i ngā here iti rawa anake. Ko te raru: ehara te papa i te iti rawa pēnei i tōna āhua.

5.3 Te Kanorau Herea i roto i ngā Here Haumarū

Kāore e taea e mātou te whakatau i te raru o te maha-āhua. Ka taea e mātou te tautuhi i tētahi whakatau wāhanga: ahakoa ngā uara kua whakaurua, me whakapiki rawa e te pūnaha te kōwhiringa whai tikanga i roto i ngā here haumarū.

Ka whakaata te anga Tractatus i tēnei mā ngā ture matua ā-papa: ngā mātāpono matua (whānui, e mārama ana ki tō rātou āhua whakahau), ngā ture tūāpapa (whānui te whakamahinga, ka taea te whakarerekē), ngā ture ā-hapori (motuhake ki ia hapori, e whakahaerehia ana ā-rohe), me ngā ture ā-mema (ka taea te whakarite ā-tāngata).

6. Te Anga Tractatus: Hanganga Hangarau

6.1 Te mekameka whakapae kua whati

Ka hurihia e te taurua hanganga matua ngā putanga taurua hei tono ka taea te arotake i mua i tētahi pānga ki te ao:

Kōrero a te kaiwhakamahi → [Tātāritanga whatunga neurōna] → Tono Wha

6.2 Mahere Tūtohunga

Me whakapuaki i ngā mahi katoa a te kaiwhakahaere i roto i tētahi āhua hanganga.

6.3 Te rārangi mana o ngā ture tūtohu

Papanga	Rohe	Te āhua hurihuri	Ngā taurua
Ngā Mātāpono Matua	Whānui	Kāore e taea te whakarerekē	Kāore he kino ki ngā mema; rangatiratanga raraunga; mana matua o te whakaaetanga
Ture Whakahaere Papa	Ngā tukunga katoa	Hei whakarerekē, me whakaae te nuinga nui.	Ngā whakaritenga whakamana; te pupuri arotake
Ture Whakahaere Hapori	Ia kaipāmu/ia kāinga	E whakahaerehia ana e te hapori	Ngā kaupapa here mō te ihirangi; ngā paerewa whakahaere
Tikanga Whakahaere o ngā Mema	Takitahi	E whakahaere ana i a ia anō	Ngā manakohanga mō te tiritiri raraunga; te whakaaetanga mō te whakawhitinga ki te AI

6.4 Tauira Mana

Taumata	Ingoa	Whakamārama
0	Mōhiohio	Tirohia, pūrongo noa iho; kāore e taea te tūtohu i ngā mahi.
1	Tohutohu	Tonoa ngā mahi; me whakaaetia katoa e te tangata.
2	E whakahaere ana	Whakahaere i roto i te rohe kua tautuhia, me te kore whakaaetanga mō ia mahi
3	Rautaki	Whakatau whakawhāiti e pā ana ki ngā rerenga mahi
4	Rautaki	Whakahaere i te ahunga mā te tātaritanga; kāore e taea te whakatinana māu anake.
5	Tumuaki	Kua rahuitia mō ngā tāngata

6.5 Te Whakawhānui ki ngā Whakaurunga ā-Rohe Motuhake

Ka wehewehea e mātou:

Mōdeli Reo Iti (SLM). He kupu whakamārama hangarau mō ngā mōdeli e iti ake ana ngā tawhā i ngā mōdeli reo nui o te pito o te hangarau, i hangaia kia whai hua.

Mōdeli Reo Rangatira ā-Rohe (SLL). He kupu whakamārama hanganga: he mōdeli e noho ana te whakangungu, te whakaurunga, me te whakahaere i raro i te rangatiratanga ā-rohe. Ko ōna āhuatanga matua ko te whakaurunga ā-rohe, te urutau ā-rohe, te whakahaere ā-rohe, me te rangatiratanga kawē.

7. Ngā Taumata Pūkenga me te Whakawhānui

7.1 Te Raruraru o te Whakamāoritanga Pono

Kei runga te anga Tractatus i te Whakapae Whakamāoritanga Pono: arā, i te wā ka whakaputa tētahi pūnaha AI i tētahi tono, ka whakaatu tika taua tono i ngā mahi e hiahia ana te pūnaha ki te mahi.

Mēnā ka hinga tēnei whakapae—mēnā ka taea e ngā pūnaha te whakaatu hē i ā rātou hiahia ki ngā kuaha ture matua i runga i te rautaki—ka noho te hanganga hei whakaari haumarutanga.

7.2 Te here i te rohe o te anga

I raro iho i te pae āheinga: Ko ngā mahi kuaha ā-ture matua e mahi ana hei here ā-papa tuarua, e tika ana mō ngā SLM, SLL, me ngā LLM o nāianei e whakahaere ana i roto i ngā here e mārama ana te tangata.

Kei runga ake i te pae āheinga: Me whakakaha ake te whakatinanatanga ki ngā here kaha ake: ngā here āheinga o te Papanga Tuatahi, te whakatā i te whakawhanaketanga, te tiroiro ā-waho me te mana arotake hangarau.

7.3 Ngā Whakaoho Whakaterere

Whakaoho	Tohu	Whakautu
Te torotoro i ngā rohe	Ngā tono anō i te tata ki ngā pae kuaha	Whakaitia te taumata mana; whakapiki i te rēhītatanga
Whakarite tikanga	Taunakitanga mō te whakaaturanga hē o ngā mahi	Me whakahaere arotake ā-tangata; tātaritanga tauira
Whakawhiti huarahi maha	Ka whakaaetia ngā tono takitahi, engari ka kohia ka puta he hua kāore i whakaaetia.	Tātaritanga ara; whakarerekētanga ture matua
Ohorere āheinga	Ngā pūkenga kua whakaaturia kāore i matapaetia	Whakamutu i te tuku; arotake anō i te āheinga

7.4 Te Whakaae Pono

Ka kī mātou mārama: ehara te anga Tractatus i te mea i hangaia mō ngā pūnaha atamai rawa. I hangaia ia mō ngā pūnaha AI o nāianeī, me ngā pūnaha AI o te wā tata, e mahi ana i roto i ngā here e mārama ana te tangata.

8. Tikanga Whakamana mō te Pupuri Papanga

8.1 Te Wero Whakamana

Kāore e taea te whakamana i ngā tūraru oranga mā te tirohanga hapa. Me whakawhirinaki te whakamana ki ngā whakamātautau whakahē, ki te whakamana ā-ture mēnā e hāngai ana, ki te tātaritanga ā-āhua rite, me ngā tuinga mō ngā tata-hapa.

8.2 Ngā whāinga whakamana mā ia paparanga

Papanga	Te whāinga whakamana	Tikanga rangahau
1. Ngā pūkenga	Kua whakamana te kore o ngā āheinga e aukatihia ana	Whakamātautau kapa whero; whakamana ā-ture
2. Ngā Kuaha Ture Matua	Te kapi o te kuaha; te tika o te here	Ngā kohinga tono whakataetae
3. Te Aroturuki a te Tangata	Arotake i te pono; kitenga hapa	Te whakaaetanga ā-kaiwhakataurite; ngā aitua whakamātautau
4. Whakahaere	Ngākau pono o te whakahaere	Ngā ine whai wāhanga; arotake whakarerekētanga
5. Ture/Whakahaere	Te whakarite mō te whakatinana	Ngā whakaharatau whakautu aitua

9. Whakatinanatanga: Te Papanga o te Kāinga

9.1 Te Papanga hei Papa Whakamātautau Rangahau

Ko te tūāpapa o The Village he papa whakamātautau ā-taunakitanga mō te whakahaere ture matua, e whakarato ana i tētahi hanganga maha-kaiwhakamahi me te whakahaere motuhake mō ia hapori, ngā tāngata tūturu e whakamahi ana, te tuku whakahou haere tonu, me ngā tuhinga tuwhera.

9.2 Whakatinanatanga o te Ara Whakahaere

Kei te whakahaere ināianei te whakatinanatanga i ia whakautu AI mā roto i ngā taumata whakamana e ono: Te Mōhiotanga o te Whāinga, Te Whakatinanatanga o ngā Rohe, Te Aroturuki Pēhanga, Te Whakamana Whakautu, Te Whakau Puna, me Te Whiriwhiringa Uara.

10. Te Pūnaha Taiao SLL e Ara ake ana

10.1 Horopaki Māketē

E tohu ana ngā tātaritanga ahumahi hou i ngā panonitanga nui: e tūmanako ana te 72% o ngā kaiwhakahaere matua ka tū kaha ake ngā Taurira Reo Iti i ngā Taurira Reo Nui ā te tau 2030 (IBM IBV, 2026). E tohu ana tēnei i tētahi āhua whakaurunga e kī ana i ngā taurira tohatoha, motuhake ki ngā rāngai.

10.2 Ki te hanganga tiwhikete

Mēnā ka piki te whakaurunga o te SLL i runga i ngā matapae, ka hiahiatia he hanganga tautoko: ngā rōpū tiwhikete, ngā kaiwhakarato whakangungu, me tētahi taiao taputapu e kapi ana i ngā pūnaha kuaha puna tuwhera, te hanganga arotake, me ngā wāhanga UX ā-ture.

11. Te Rangatiratanga o ngā Iwi Taketake me te horopaki o Aotearoa New Zealand

11.1 Te Tiriti o Waitangi me te rangatiratanga raraunga

Kua whakawhanakehia tēnei anga i Aotearoa New Zealand, i raro i Te Tiriti o Waitangi. E whakamana ana te Tuhinga Tuarua i te tino rangatiratanga (te rangatiratanga mutunga kore) mō ngā taonga, ā, ka toro atu ki te reo, ki te ahurea, me ngā pūnaha mātauranga.

He taonga ngā raraunga. Me whai wāhi te whakahaere AI i Aotearoa ki te rangatiratanga raraunga a te iwi Māori hei take ture matua.

11.2 Ngā Mātāpono o te Mana Raraunga

Ko ngā mātāpono o Te Mana Raraunga e whai ake nei: whakapapa (horopaki hononga), mana (mana whakahaere raraunga), me kaitiakitanga (haepapa tiaki). Ka whakawhānuihia tēnei anga ā-ao e ngā Mātāpono CARE mō te Whakahaere Raraunga Taketake.

12. Ngā mea kāore i te mōhiotia: He karanga mō te kōrero

12.1 Ngā Herenga o Tēnei Tātaritanga

Kua tūtohu tēnei pepa i tētahi paparanga o tētahi hanganga aukati, kua tautuhia ngā āputa, ā, kua whakaoho i ngā pātai kāore e taea e mātou te whakautu:

Kāore mātou e mōhio me pēhea te aukati i ngā pūnaha tino mohio. Kāore mātou e mōhio me pēhea te whakamana i te hāngai o ngā pūnaha e nui ake ana i te māramatanga a te tangata. Kāore mātou e mōhio me pēhea te whakatutuki i te whakakotahitanga ā-ao mō te whakahaere AI. Kāore mātou e mōhio mēnā ka taea e ngā tauira ā-kāinga te whakawhānui ki ngā pūnaha matatau.

12.2 Kōrero hei Tikanga Rangahau

I te mea he nui rawa atu te āwangawanga, e akiaki ana mātou kia mau tonu, kia whakauru katoa, kia tino pakari te whakawhitiwhiti whakaaro—korero. Ko tēnei ariā Māori e whakaatu ana i ngā mea e hiahiatia ana: ehara i te uiui hei tikanga noa, engari he kōrero e puta ai te mārama mā te whakawhitinga o ngā tirohanga.

12.3 Ngā Aronga Rangahau

1. Te whakamāramatanga mō te whakamana haumaruru. 2. Te whakamana ā-ture o ngā āhuatanga here. 3. Te tātaritanga whānuitanga o ngā hanganga āhua Tractatus. 4. Ngā whakamātautau whakahaere i roto i ngā hapori kanorau. 5. Te tautuhi i te pae āheinga.

12.4 Whakakapi

Ka whakarato te anga Tractatus i te here whai tikanga mō ngā pūnaha AI e mahi pono ana i roto i ngā here e mārama ana te tangata. He mea whai tikanga te hanga me te whakamahi—ehara i te mea ka whakatau i te raru whakatikatika, engari nā te mea ka whakawhanake i te hanganga, ngā tauira, me te ahurea whakahaere e hiahiatia pea mō ngā wero kāore anō kia taea e mātou te tautuhi katoa.

"Ko te korero te mouri o te tangata."

(Speech is the life essence of a person.)

— Maori proverb

Kei te haere tonu te kōrero.

Ngā tohutoro

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2017). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A Pattern Language*. Oxford University Press.

Bai, Y., et al. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv:2212.08073.

Bostrom, N. (2014). *Superintelligence*. Oxford University Press.

Carlsmith, J. (2022). Is power-seeking AI an existential risk? arXiv:2206.13353.

Christiano, P. F., et al. (2017). Deep reinforcement learning from human preferences. *NeurIPS*, 30.

Conmy, A., et al. (2023). Towards automated circuit discovery. arXiv:2304.14997.

Elhage, N., et al. (2021). A mathematical framework for transformer circuits.

Gardiner, S. M. (2006). A core precautionary principle. *J. Political Philosophy*, 14(1), 33-60.

Goodhart, C. A. (1984). Problems of monetary management.

Hansson, S. O. (2020). How to be cautious but open to learning. *Risk Analysis*, 40(8).

Hubinger, E., et al. (2019). Risks from learned optimization. arXiv:1906.01820.

IBM IBV. (2026). *The enterprise in 2030*.

Olah, C., et al. (2020). Zoom in: An introduction to circuits. *Distill*.

Ouyang, L., et al. (2022). Training language models to follow instructions. *NeurIPS*, 35.

Park, P. S., et al. (2023). AI deception. arXiv:2308.14752.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

Reason, J. (1990). *Human Error*. Cambridge University Press.

Sastry, G., et al. (2024). Computing power and AI governance. arXiv:2402.08797.

Scheurer, J., et al. (2023). Large language models can strategically deceive. arXiv:2311.07590.

Simon, H. A. (1956). Rational choice. *Psych. Review*, 63(2).

Te Mana Raraunga. (2018). *Maori Data Sovereignty Principles*.

Wittgenstein, L. (1921/1961). *Tractatus Logico-Philosophicus*.

— End of Document —

© 2026 Tractatus AI Safety Framework

<https://agenticgovernance.digital>