



Pagsusuri sa bisa ng *AI content detectors* sa pagtukoy ng orihinalidad ng akademikong sulatin sa Filipino

Ma. Cara Tanya B. Princillo*¹, Brian S. Delopere², Wayne David C. Padullon³

^{1, 3} Leyte Normal University, Tacloban City, Philippines, ² Dulag National High School, Leyte, Philippines

Corresponding Author email: macaratanya.princillo@lnu.edu.ph

Received: 06 December 2025

Revised: 14 April 2026

Accepted: 01 May 2026

Available Online: 05 May 2026

Volume 1 (2026), Issue 2, P-ISSN – 3116-4463; E-ISSN - 3116-4471

<https://doi.org/10.63498/pdpf3>

Abstrak

Layunin: Nilalayon ng pag-aaral na ito na suriin ang kabisaan at kahusayan ng anim na AI content detectors—GPTZero, Smodin.Ai, Isgen.Ai, ZeroGPT, Justdone, at OpenL.io—sa pagtukoy ng orihinalidad ng mga akademikong sulatin na isinulat sa wikang Filipino, partikular sa pagkilala ng AI-generated at human-written na mga teksto.

Metodolohiya: Gumamit ang pag-aaral ng explanatory sequential mixed-methods approach. Sa kwantitatibong yugto, isinagawa ang estadistikal na pagsusuri gamit ang precision, recall, at F1 score upang masukat ang performance ng bawat AI detector. Sa kwalitatibong yugto, isinagawa ang masusing panayam sa mga mag-aaral mula sa programang BSEd-Filipino upang suriin ang kanilang mga karanasan, pananaw, at saloobin hinggil sa paggamit ng AI content detection tools.

Resulta: Ipinakita ng mga resulta na may makabuluhang pagkakaiba ang mga AI content detectors sa kanilang antas ng precision, recall, at F1 score. Natuklasan din na may mataas na insidente ng false positives at indikasyon ng bias ang ilang tool, lalo na kapag sinusuri ang mga tekstong nakasulat sa wikang Filipino. Mula sa kwalitatibong datos, lumitaw na ang AI detection results ay may epekto sa emosyon, tiwala sa sarili, at motibasyon ng mga mag-aaral sa akademikong pagsulat.

Kongklusyon: Ipinahihiwatig ng mga natuklasan na hindi sapat ang AI content detectors bilang nag-iisang batayan sa pagsusuri ng orihinalidad ng akademikong sulatin. Iminumungkahi ng pag-aaral ang mas responsableng paggamit ng mga AI detection tools na sinusuportahan ng human judgment at ang pangangailangan para sa pagbuo ng mas inklusibo at kontekstual na mga sistema na sensitibo sa wikang Filipino at sa kontekstong edukasyonal ng Pilipinas.

Susing-salita: *AI content detectors, orihinalidad, akademikong integridad, wikang Filipino, false positives, F1 score*

INTRODUKSYON

Sa patuloy na pag-unlad ng teknolohiya sa pandaigdigang konteksto, lalong lumalawak ang impluwensiya ng Artificial Intelligence (AI) sa iba't ibang larangan, kabilang ang edukasyon. Isa sa mga pangunahing gamit ng AI sa akademikong mundo ay ang AI content detectors, na idinisenyo upang tukuyin ang antas ng orihinalidad ng isang sulatin at matukoy kung ito ay gawa ng tao o binuo ng awtonomong teknolohiya. Gayunpaman, sa kabila ng kanilang mabilis na pag-usbong, lumitaw ang mga isyu kaugnay ng kawastuhan at pagiging maaasahan ng mga sistemang ito (Sadasivan et al., 2023). Ayon kina Weber-Wulff et al. (2023), maraming AI content detector ang nahihirapang kilalanin ang mga teksto na binuo ng mas sopistikadong AI models na maaaring makalampas sa mga mekanismo ng deteksiyon. Bukod dito, itinuturo ng ilang pag-aaral na maaaring magkaroon ng false positives, mga pagkakataong ang orihinal na sulatin ng isang mag-aaral ay pinaghihinalaang gawa ng AI, na maaaring magdulot ng emosyonal, sikolohikal, at akademikong pinsala (Hirsch, 2024; Abdali et al., 2024).

Sa antas ng pambansang diskurso, binibigyang-pansin ng mga edukador sa Pilipinas ang mga implikasyon ng paggamit ng AI sa pagpapatili ng integridad ng mga akademikong gawain. Sa antas ng patakarang pang-edukasyon, binibigyang-diin ng mga institusyong pang-akademiko sa Pilipinas ang kahalagahan ng integrasyon ng AI literacy sa kurikulum upang maturuan ang mga mag-aaral sa responsableng paggamit ng teknolohiya at upang maiwasan ang maling paggamit nito sa paggawa ng mga akademikong sulatin (Holmes et al., 2021). Gayunpaman, lumilitaw ang mga suliranin sa ilang AI content detectors, partikular ang pagkiling laban sa mga manunulat na gumagamit ng wikang di-Ingles, gaya ng Filipino. Ipinunto ni Coldwell (2024) na mas mataas ang posibilidad ng false



positive sa mga AI detector kapag ang teksto ay isinulat sa ibang wika bukod sa Ingles, isang usapin na may malalim na epekto sa mga bansang multilingguwal tulad ng Pilipinas, kung saan napapailalim sa posibleng diskriminasyon ang mga akademikong sulatin na nasa wikang Filipino.

Sa lokal na konteksto, patuloy na hinaharap ng mga guro ng Filipino ang hamon ng pagpapanatili ng akademikong integridad sa gitna ng digital na transisyon. Ayon kina Montalan et al. (2025), bagamat ang mga AI content detector ay inobatibong kasangkapan para sa pagsusuri ng mga sulatin, lumilitaw ang pangangailangan para sa mas malalim na pagsusuri ng kanilang pagiging epektibo sa wikang Filipino. Ipinapakita ng ulat ng Palmer (2024) na may mga pagkakataon na ang mga AI detector ay nagbibigay ng hindi akmang resulta kapag sinuri ang mga sulatin sa Filipino, bagay na maaaring makaapekto sa kredibilidad ng mga guro at tiwala ng mga mag-aaral. Dagdag pa rito, binigyang-diin ni Giray (2024) na may kakulangan ang kasalukuyang mga teknolohiya sa pagkilala ng malikhaing pagpapahayag sa mga lokal na wika gaya ng Filipino, dahil karamihan sa mga ito ay nakasentro sa estruktura at pamantayan ng wikang Ingles.

Naglalayon ang pananaliksik na ito na suriin ang kabisaan ng mga AI content detector sa pagtukoy ng orihinalidad ng mga akademikong sulatin na nakasulat sa wikang Filipino. Ayon kina Estrellado at Miranda (2023), mahalaga ang masusing pagsusuri sa etikal na paggamit ng AI sa edukasyon upang matiyak ang integridad ng akademikong sulatin sa wikang Filipino. Magsisilbi itong mahalagang ambag sa diskurso ng AI ethics sa edukasyon, at maaaring magsilbing batayan para sa mga guro, mag-aaral, at institusyon sa pagbuo ng mas makatarungan at makataong patakaran hinggil sa paggamit ng teknolohiya (Holmes et al., 2021). Higit pa rito, inaasahang mapalalalim ng pananaliksik ang pang-unawa sa kung paano maaring gamitin ang AI sa paraang hindi lamang teknolohikal kundi makatao, maka-Filipino, at etikal, alinsunod sa mga rekomendasyon nina Fabro et al. (2024) na nagsusulong ng AI literacy at responsableng paggamit ng AI sa akademikong konteksto.

Ang pag-aaral ay ginabayan ng mga konsepto ng akademikong integridad, AI ethics sa edukasyon, at kolaborasyon ng tao at teknolohiya, na nagbibigay-diin sa papel ng human judgment bilang mahalagang salik sa pagsusuri ng orihinalidad ng akademikong sulatin. Ang mga konseptong ito ang nagsilbing batayan sa pagsusuri ng bisa ng AI content detection tools at sa interpretasyon ng mga karanasan ng mga mag-aaral sa kontekstong lingguwistiko ng Filipino.

Sa kabila ng lumalawak na pananaliksik sa bisa ng AI content detectors sa internasyonal na konteksto, nananatiling limitado ang empirikal na ebidensiya hinggil sa kanilang pagganap kapag inilalapat sa mga akademikong sulatin na nakasulat sa wikang Filipino. Karamihan sa mga umiiral na pag-aaral ay nakatuon sa mga tekstong Ingles at bihirang isaalang-alang ang impluwensiya ng istruktura, diskurso, at lingguwistikong katangian ng Filipino sa katumpakan ng AI detection. Dahil dito, may puwang sa kaalaman ukol sa pagiging epektibo, patas, at etikal ng mga AI content detection tools sa lokal na kontekstong Pilipino, na siyang tinutugunan ng kasalukuyang pag-aaral.

Ang pananaliksik na ito ay nag-aambag sa larangan ng edukasyon sa pamamagitan ng dalawang mahahalagang aspekto. Una, empirikal nitong sinusuri ang pagganap ng anim na AI content detection tools gamit ang mga sukatan ng katiyakan (precision), pagkilala (recall), at pinagsamang sukatan (F1 score) sa mga sanaysay na nakasulat sa wikang Filipino. Ikalawa, pinagsasama nito ang kwantitatibong pagsusuri ng detection accuracy at ang kwalitatibong pagsusuri ng karanasan at pananaw ng mga mag-aaral, na nagbibigay ng mas komprehensibo at makataong pag-unawa sa implikasyon ng AI detection sa akademikong pagsulat sa Filipino.

Rebyu ng Kaugnay na Literatura at Pag-aaral

Potensiyal at Hamon ng AI sa Akademikong Pagsulat

Sa kasalukuyang kalagayan ng edukasyon, laganap ang paggamit ng *Artificial Intelligence (AI)* bilang instrumento sa pag-aaral ng mga mag-aaral. Kaugnay nito, sa pag-aaral nina Dingal et al. (2024), tinalakay ang epekto ng paggamit ng AI sa akademikong pagsulat ng mga mag-aaral. Ipinahayag ng mga resulta na may makulay na relasyon ang paggamit ng AI at ang kalidad ng pagsulat ng mga mag-aaral, partikular sa mga aspeto ng *AI literacy*, tiwala, at kaugnyanan ng AI. Nagbigay ng positibong epekto sa pagkatuto ng mga mag-aaral at mag-ambag sa pagbuo ng maunlad na teknolohiya at kolaboratibong kapaligiran sa edukasyon.

Samantala, binigyang-diin naman sa pag-aaral nina Ordonez et al. (2024) na ang integrasyon ng AI ng mga guro ay nagdulot ng kaghinawaan, pagiging maaasahan, at kasiyahan sa pagtuturo, partikular sa pagbuo ng mga kagamitan sa pagtuturo at pagsusuri sa mga mag-aaral. Bukod pa rito, ipinakita na ang mas mataas na pagtingin ng mga mag-aaral sa kakayahan at kapakinabangan ng *large language models* (LLMs), tulad ng ChatGPT, ay nagdulot ng pagpapabuti sa kritikal na pag-iisip ($r > .583$) at *self-efficacy* ($r > .603$), na nagpapahiwatig ng positibong ugnayan sa pagitan ng paggamit ng AI at pag-unlad sa mga kasanayang pang-akademiko.



Sa kabilang banda, sa pag-aaral nina Pecasion et al. (2023) ay layunin ang pagsusuri ng pananaw ng mga guro sa edukasyon sa paggamit ng AI, tulad ng ChatGPT, bilang isang kasangkapan sa kanilang pang-araw-araw na trabaho. Napag-alaman sa pag-aaral na ang ChatGPT ay naging mahalagang kasangkapan sa mga guro, na nagpapadali at nagpapakomportable ng mga gawain at tungkulin. Bagamat nakikitang kapaki-pakinabang ang teknolohiyang ito, itinuro ng mga guro na kailangang mag-ingat at may tamang etika at kaugalian sa paggamit nito.

Higit pa rito, sa isang sarbey na isinagawa ni Nam (2023), 56% ng mga mag-aaral sa kolehiyo ang umamin na gumamit ng AI sa kanilang mga asignatura, akademikong pagsusulat o pagsusulit. Habang hindi lahat ng paggamit na ito ay itinuturing na pandaraya, ipinapakita ang lumalawak na pagdepende ng mga mag-aaral sa AI sa pag-aaral.

Katumpakan, Limitasyon, at Bias ng AI Content Detectors

Ayon kay Gao et al. (2022), isang posibleng solusyon sa isyu ng pagtukoy ng *AI-generated* na teksto ay ang paggamit ng *AI content detectors* na may kakayahang magsuri at matukoy kung ang teksto ay nilikha ng isang sistema ng AI. Sa kanilang pagsusuri ay lumitaw na bagamat may kakayahan ang mga ito na tukuyin ang mga *AI-generated* na nilalaman, may mga limitasyon pa rin sa antas ng katumpakan ng mga ito.

Kaugnay nito, ipinakita sa pag-aaral nina Weber-Wulff et al. (2023) na nagkakaroon ng mga *false positives* at *false negatives* sa pagtukoy ng *AI-generated* na sanaysay, na maaaring humantong sa hindi patas na pagsusuri sa mga gawa ng mag-aaral. Partikular, lumitaw na ang mga libreng *AI-detection tools* ay may pabagu-bagong *predictive value*, na nangangahulugang hindi palaging maaasahan ang mga resulta. Ang pagiging sensitibo ng mga tool na ito ay maaaring magbago sa paglipas ng panahon kasabay ng mga *pag-upgrade* ng *software*.

Samantala, sa pananaliksik nina Angeles et al. (2024), sinuri ang pananaw ng mga mag-aaral hinggil sa paggamit ng *AI Detector System* ng kanilang mga guro sa pagtukoy ng orihinalidad ng mga akademikong sulatin. Bagamat kinikilala ang kahalagahan ng mga tool na ito sa pagpapanatili ng integridad, nabanggit din ang mga alalahanin tulad ng maling positibong resulta at kakulangan ng paliwanag kung paano sinusuri ng AI ang kanilang mga gawa.

Dagdag pa rito, ipinunto ni Wu et al. (2023) na ang epektibo o tiyak na kakayahan ng mga *AI content detector* sa pagtukoy ng mga sanaysay ng mag-aaral ay nananatiling hindi pa lubos na tiyak. Halimbawa, kung gagamit ang mag-aaral ng AI text generator at pagkatapos ay i-edit ito sa pamamagitan ng paraphrasing o pagbabago ng estruktura ng pangungusap, maaaring malinlang ang mga *AI detector*. Ipinapakita nito ang kahinaan ng mga detector sa paghawak ng mga manipuladong *AI-generated content*.

Sa kabilang banda, binigyang-diin nina Liang et al. (2023) na may AI detectors na nagpapakita ng *bias* laban sa mga hindi katutubong nagsasalita ng Ingles. Ipinakita sa kanilang pag-aaral na mas mataas ang insidente ng *false positives* sa mga sulatin ng mga mag-aaral na may limitadong kasanayan sa Ingles. Higit pa rito, napatunayan nilang ang mga simpleng estratehiya sa *pag-prompt* ay maaaring malampasan ang mga GPT detector, na nagpapakita ng kahinaan ng mga ito sa pantay-pantay na pagtrato sa mga manunulat na may iba't ibang antas ng kasanayan sa wika. Dahil dito, mahalagang magsagawa ng pagsisiyasat upang mabigyan ng gabay ang mga mananaliksik at akademiko sa pagpili ng angkop na *software* para sa kanilang pangangailangan sa pagsusulat.

Kolaborasyon ng Human-AI at Mga Implikasyon sa Patakarang Pang-edukasyon

Sa pananaliksik nina Giray et al. (2025), inilahad na ang mga *AI content detector* ay madalas na hindi epektibo kapag ang isang sanaysay ay may mataas na antas ng kolaborasyon sa pagitan ng tao at AI, kung saan ang mag-aaral ay aktibong nag-aambag ng sariling ideya habang ginagamit ang AI bilang gabay sa estruktura at pagpili ng mga salita. Batay sa kanilang natuklasan, nahihirapan ang mga *AI detector* sa pagtukoy ng ganitong uri ng pinagsamang gawain, kaya't lumilitaw ang mga isyu sa katumpakan ng pagsusuri. Sa kabilang banda, nakita rin nila na ang kolaboratibong paggamit ng AI sa pagsusulat ay nagbubunga ng mas positibong resulta sa pagkatuto at aktibong partisipasyon ng mga mag-aaral. Dahil dito, binigyang-diin ng kanilang pag-aaral ang kahalagahan ng pagkakaroon ng malinaw na gabay mula sa mga institusyon ukol sa tamang paggamit ng AI at *AI detection*, gayundin ang pangangailangang sanayin ang mga guro at mag-aaral sa mga proseso ng pagsusulat na hindi lamang nakatuon sa produkto kundi sa buong proseso.

Ipinakita sa mga lokal na pag-aaral at ulat ang mabilis na tugon ng mga institusyong pang-edukasyon sa Pilipinas sa pag-usbong ng AI. Sa ulat ng University of the Philippines Open University (2023), binuo ng institusyon ang *AI Guidelines* na nagbibigay-diin sa etikal na paggamit ng generative AI sa mga gawaing akademiko. Nilinaw ng patakaran na maaaring gamitin ang AI bilang kasangkapan sa pagbuo ng ideya, balangkas, at teknikal na aspeto ng pagsusulat, ngunit dapat manatili ang aktibong partisipasyon ng tao sa pagsusuri at paglikha ng nilalaman.



Samantala, ang De La Salle University (DLSU) ay nagpahayag din ng kanilang paninindigan sa pamamagitan ng pagbuo ng institutional policy na nagbibigay-gabay sa paggamit ng AI sa pagtuturo, pananaliksik, at mga gawaing akademiko. Sa kanilang inilabas na panuntunan, binibigyang-diin ang responsableng paggamit ng AI na hindi sumasagka sa orihinalidad, integridad, at malayang pag-iisip ng mag-aaral (De La Salle University, 2023). Ipinapakita ng mga patakarang ito na ang mga unibersidad sa Pilipinas ay nagsisimula nang lumikha ng mga konkretong hakbang upang mapanatili ang balanse sa pagitan ng paggamit ng teknolohiya at paghubog ng kritikal na pag-iisip, isang mahalagang aspeto ng kolaborasyon sa pagitan ng tao at AI sa edukasyong Pilipino.

Sa kabuoan, pinatutunayan ng mga pag-aaral na habang makabuluhan ang papel ng AI sa pagbuo ng akademikong sulatin, may mga umiiral na limitasyon ang mga *AI content detector*, lalo na sa pagkilala ng komplikadong *AI-generated texts* at pagsusuri sa iba't ibang wika. Bunsod nito, nagiging mas mahalaga ang tuloy-tuloy na pananaliksik upang masuri kung paano maisasama ang AI sa edukasyon nang hindi nasasakripisyo ang akademikong integridad. Sa gayon, kinakailangang patuloy pang pag-aralan ang epektibong integrasyon ng AI sa edukasyon upang mapanatili ang balanse sa pagitan ng teknolohikal na pag-unlad at akademikong integridad.

Teoretikal na Balangkas

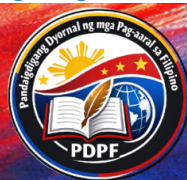
Sa kabila ng lumalawak na pananaliksik hinggil sa AI content detection tools sa pandaigdigang konteksto, nananatiling limitado ang mga empirikal na pag-aaral na sumusuri sa kanilang bisa sa mga akademikong sulatin na nakasulat sa wikang Filipino. Karamihan sa mga umiiral na pag-aaral ay nakatuon sa wikang Ingles, kung kaya't hindi pa ganap na nauunawaan ang pagganap ng mga AI detector sa mga kontekstong multilingguwal tulad ng Pilipinas (Liang et al., 2023). Dahil dito, mahalagang iangkla ang kasalukuyang pag-aaral sa mga kinikilalang teorya upang masuri hindi lamang ang teknikal na katumpakan ng mga tool kundi pati ang kanilang implikasyon sa pagsusuri at paggamit sa edukasyon.

Ang pag-aaral na ito ay nakabatay sa dalawang pangunahing teorya: Sociotechnical Systems Theory at Validity Theory in Educational Measurement.

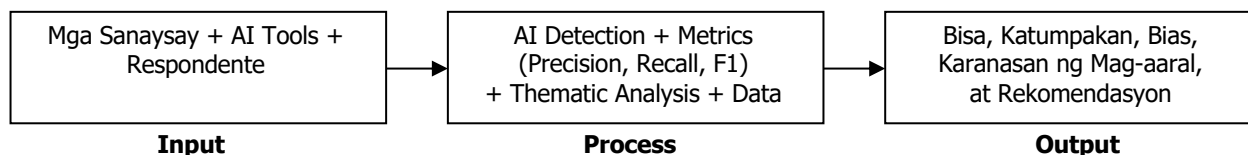
Una, ang Sociotechnical Systems Theory nina Trist at Bamforth (1951) ay nagsasaad na ang teknolohiya at ang mga taong gumagamit nito ay bumubuo ng isang magkakaugnay na sistema, kung saan ang bisa ng teknolohiya ay hindi lamang nakabatay sa disenyo nito kundi sa paraan ng paggamit at interpretasyon nito ng tao. Sa konteksto ng pag-aaral na ito, ang AI content detection tools ay hindi dapat ituring bilang hiwalay at awtomatikong tagapagpasya ng orihinalidad ng akademikong sulatin. Sa halip, ito ay bahagi ng mas malawak na sistemang kinabibilangan ng guro, mag-aaral, at institusyong pang-edukasyon. Ipinapahiwatig ng teoryang ito na ang resulta mula sa AI detectors ay nangangailangan ng masusing *human judgment*, sapagkat ang labis na pag-asa sa teknolohiya nang walang interpretasyon ng tao ay maaaring magdulot ng maling desisyon, lalo na sa mga kasong may *false positives* o maling klasipikasyon.

Ikalawa, ang Validity Theory in Educational Measurement, partikular ang pananaw ni Messick (1989), ay nagbibigay-diin na ang bisa (*validity*) ng isang instrumento ay nakabatay hindi lamang sa katumpakan ng sukat kundi pati sa kahulugan at implikasyon ng interpretasyon ng mga resulta. Ayon kay Messick, ang validity ay sumasaklaw sa parehong teknikal na aspeto (accuracy at consistency) at sa mga kahihinatnang panlipunan ng paggamit ng isang instrumento. Sa pag-aaral na ito, ang AI content detection tools ay maituturing na mga instrumentong pangpagsusuri na sinusukat gamit ang mga sukatan tulad ng precision, recall, at F1 score. Gayunpaman, hindi sapat na mataas ang mga numerikal na sukatan upang masabing epektibo ang isang tool. Mahalaga ring suriin ang mga implikasyon ng kanilang paggamit, gaya ng posibilidad ng bias laban sa wikang Filipino at ang epekto ng maling klasipikasyon sa karanasan, motibasyon, at tiwala sa sarili ng mga mag-aaral.

Sa kabuuan, ang pagsasanib ng Sociotechnical Systems Theory at Validity Theory in Educational Measurement ay nagbibigay ng komprehensibong teoretikal na lente sa pag-aaral na ito. Pinagsasama nito ang pag-unawa sa ugnayan ng teknolohiya at tao sa proseso ng pagsusuri, at ang pagsusuri sa bisa at implikasyon ng mga AI detection tools bilang mga instrumentong pang-edukasyon. Sa ganitong paraan, natutugunan ng pag-aaral ang umiiral na research gap sa konteksto ng wikang Filipino at nag-aambag sa mas makatarungan, maingat, at kontekstual na paggamit ng AI sa edukasyon.



Balangkas Konseptwal



Ang konseptwal na balangkas ng pag-aaral na ito ay nakabatay sa Input–Process–Output (IPO) Model, na nagpapakita ng sistematikong daloy ng pagsusuri sa bisa ng AI content detection tools sa akademikong sulatin sa wikang Filipino. Sa input, kabilang ang mga sanaysay (AI-generated at human-written), mga AI detection tools, at mga respondente. Sa process, isinagawa ang kwantitatibong pagsusuri gamit ang precision, recall, at F1 score kasama ang confusion matrix, gayundin ang kwalitatibong pagsusuri sa pamamagitan ng panayam at thematic analysis, na kalauna'y isinama upang mapalalim ang interpretasyon ng datos. Sa output, natukoy ang antas ng bisa at katumpakan ng mga AI tools, ang mga limitasyon at posibleng bias nito sa wikang Filipino, pati ang mga epekto nito sa karanasan ng mga mag-aaral, na nagsilbing batayan sa pagbuo ng mga implikasyong pang-edukasyon at rekomendasyon para sa mas makatarungan at responsableng paggamit ng AI sa akademikong konteksto.

Layunin ng Pananaliksik

Pangkalahatang Layunin

Layunin ng pag-aaral na ito na suriin ang pagiging epektibo ng mga AI content detection tools sa pagtukoy ng mga akademikong sanaysay na binuo ng AI at ng mga sanaysay na isinulat ng tao sa wikang Filipino.

Tiyak na mga Layunin

Nilalayan ng pananaliksik na ito na:

1. Suriin ang pagganap ng mga *AI content detection tools* gamit ang mga sukatan ng katiyakan (*precision*), pagkilala (*recall*), at pinagsamang sukatan (*F1 score*).
2. Suriin ang katumpakan ng klasipikasyon ng mga *AI content detection tools* batay sa mga sumusunod na sukatan: *true positives*, *true negatives*, *false positives*, at *false negatives*.
3. Tukuyin at suriin ang mga pananaw, karanasan, at saloobin ng mga mag-aaral ng BSEd-Filipino hinggil sa pagiging epektibo at patas ng mga *AI content detection tools* sa pagsusuri ng akademikong sulatin.

Mga Katanungan sa Pananaliksik

Nilalayan ng pag-aaral na ito na sagutin ang mga sumusunod na katanungan:

1. Gaano kaepektibo ang mga *AI content detection tools* sa pagtukoy ng mga akademikong sanaysay na binuo ng AI at ng mga sanaysay na isinulat ng tao sa wikang Filipino batay sa katiyakan (*precision*), pagkilala (*recall*), at *F1 score*?
2. Gaano katumpak na naiuri ng mga *AI content detection tools* ang mga akademikong tekstong Filipino batay sa *true positives*, *true negatives*, *false positives*, at *false negatives*?
3. Ano ang mga pananaw, karanasan, at saloobin ng mga mag-aaral ng BSEd-Filipino hinggil sa pagiging epektibo at patas ng mga *AI content detection tools* sa pagsusuri ng akademikong pagsulat?

METODOLOHIYA

Disenyo ng Pananaliksik

Sa pag-aaral na ito, ginamit ang *explanatory sequential* na pinaghalong disenyo kwantitatibo at kwalitatibo sa pangangalap ng datos. Unang tinipon at sinuri ang kwantitatibong datos at sinundan ito ng kwalitatibong pagsusuri upang mas higit na maunawaan ang mga resulta. Mahalaga ang disenyo ito upang masukat ang kabisaan ng *AI content detection tools* at mabigyang-paliwanag kung paano ito nakaapekto sa pagsusuri ng orihinalidad ng akademikong sulatin ng mga mag-aaral (Creswell & Creswell, 2018).

Sa kwantitatibong yugto ng pananaliksik, ang mga mag-aaral mula sa BSEd-Filipino ay sumulat ng isang sanaysay hinggil sa Kahalagahan ng Wika at Panitikan. Gamit ang mga *AI content detection tools*, sinuri ang mga sanaysay na isinulat ng mga mag-aaral upang matukoy kung tama ang pagkakaklasipika ng *AI-generated* at *human-*



written na sulatin. Sinukat ang bisa ng *AI content detection tools* gamit ang katiyakan (*precision*), pagkilala (*recall*), at pinagsamang sukatan (*F1 score*) sa pamamagitan ng pagsusuring estadistika tulad ng *Chi-Square* at *ANOVA*.

Sa kwalitatibong yugto naman, isinagawa ang masusing panayam (*in-depth interview*) sa mga mag-aaral mula sa BSEd-Filipino at siniyasat ang kanilang mga pananaw, karanasan, at saloobin hinggil sa pagiging epektibo at patas ng *AI content detection tools* sa pagtatasa ng kanilang mga akademikong sulatin, gayundin ang mga salik na maaaring nakaapekto sa kanilang paraan ng pagsulat at paggamit ng wikang Filipino.

Sa pamamagitan ng *explanatory sequential* na disenyo ng pananaliksik, naunawaan ang *statistical accuracy* ng *AI content detection tools*, maging ang kontekstwal na epekto nito sa larangan ng edukasyon, batay sa pananaw, saloobin, at karanasan ng mga mag-aaral. Nagpayaman ito sa pag-unawa ng epekto ng *AI* sa larangan ng pagsulat ng mga akademikong sulatin sa Filipino.

Populasyon at Sampol

Ang mga respondente ng pag-aaral ay binubuo ng walumpung (80) mag-aaral mula sa programang BSEd-Filipino para sa kwantitatibong yugto, at sampung (10) piling mag-aaral para sa kwalitatibong yugto.

Sa kwantitatibong bahagi, ginamit ang stratified random sampling (Cochran, 1977) upang matiyak ang pantay na representasyon ng bawat antas ng pag-aaral. Ang mga kalahok ay pinili batay sa sumusunod na pamantayan: (1) kasalukuyang naka-enroll sa BSEd-Filipino; (2) may kasanayan sa pagsulat sa wikang Filipino; at (3) handang lumahok sa pagsulat ng sanaysay sa itinakdang oras.

Sa kwalitatibong bahagi, ginamit ang purposive sampling upang piliin ang mga kalahok na nakaranas ng false positive o false negative na resulta mula sa *AI detection tools*. Ang pamamaraang ito ay angkop upang makuha ang mas malalim at makabuluhang karanasan ng mga respondente kaugnay ng layunin ng pag-aaral.

Mga Ginamit na Instrumento sa Pananaliksik

A. Kwantitatibong Yugto

Ang sumusunod na instrumento ang ginamit ng mga mananaliksik sa kwantitatibong yugto ng pangangalap ng datos.

Pagsulat ng Sanaysay at *AI Content*

Sa loob ng isang (1) oras, ang mga mag-aaral mula sa BSEd Filipino ay pinasulat ng isang sanaysay na kinakitaan ng 300 - 500 na salita na pumapaksa hinggil sa "Kahalagahan ng Wika at Panitikan". Ang mga sanaysay na ito ay hinati sa dalawang kategorya: *human-written*, na isinulat nang walang tulong ng *AI*, at *AI-Generated*, na awtomatikong binuo ng isang *AI model*. Ang mga sanaysay na ito ang nagsilbing pangunahing *dataset* na pinroseso gamit ang mga *AI detection tool*.

AI Content Detection Tools

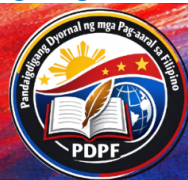
Ang kabisaan ng mga *AI content detection tool* sa tamang pagtukoy ng tekstong isinulat ng tao at ng *AI* ay sistematikong sinuri sa pag-aaral na ito. Ang mga napiling *detection tool* ay kinabibilangan ng GPTZero, Smodin.Ai, Isgen.Ai, Zero GPT, Justdone, at OpenLio. Ang bawat *tool* ay ginamit upang suriin ang parehong sanaysay na isinulat ng mga mag-aaral at ng *AI*, at ang kanilang mga resulta ng klasipikasyon ay itinala at sinuri gamit ang mahahalagang sukatan ng pagganap.

B. Kwalitatibong Yugto

Sa kwalitatibong yugto, ginamit ang isang instrumento upang mapalalim ang pag-unawa sa karanasan ng mga mag-aaral sa paggamit ng *AI detection tools*. Bago isagawa ang pangangalap ng datos, ang semi-structured interview guide ay sumailalim sa content validation ng tatlong (3) eksperto sa larangan ng wika, pananaliksik, at edukasyon. Ang kanilang mga mungkahi ay ginamit upang linawin ang mga tanong, tiyakin ang kaangkupan ng wika, at mapahusay ang kakayahan ng instrumento na makuha ang aktwal na karanasan ng mga mag-aaral. Ang *AI content detection tools* naman ay piniling batay sa kanilang malawak na paggamit at accessibility sa akademikong konteksto.

Semi-Structured Questionnaire

Ang semi-structured questionnaire na ginamit sa masusing panayam (*in-depth interview*) ay isang researcher-made instrument na binuo batay sa layunin at mga tanong sa pananaliksik. Ito ay idinisenyo upang sistematikong makuha ang pananaw, karanasan, at saloobin ng mga piling mag-aaral na nakaranas ng false positive o false negative na resulta mula sa *AI content detection tools*, kung saan ang mga tanong ay inangkop mula sa mga pangunahing baryabol ng pag-aaral, kabilang ang reaksyon sa *AI detection results*, persepsyon sa bisa at pagiging patas ng mga tool, at ang epekto nito sa kanilang akademikong pagsulat sa wikang Filipino. Upang matiyak ang kaangkupan at kalinawan ng instrumento, ito ay sumailalim sa content validation ng tatlong (3) eksperto sa larangan



ng wika, pananaliksik, at edukasyon, at ang kanilang mga mungkahi ay isinama upang mapabuti ang estruktura at linaw ng mga tanong. Ang semi-structured na format ay pinili upang magbigay ng balanse sa pagitan ng gabay na tanong at kalayaan ng mga kalahok na maglahad ng mas malalim at kontekstuwal na karanasan.

Pagkalap ng Datos

Sa kwantitatibong yugto ng pananaliksik, isinagawa ang pangangalap ng datos sa sunod-sunod na hakbang. Una, kinilala at pinili ang mga respondente na binubuo ng walumpung (80) mag-aaral mula sa iba't ibang antas ng BSEd-Filipino na may kasanayan sa akademikong pagsulat gamit ang wikang Filipino. Ikalawa, pinasulat ang mga respondente ng isang sanaysay hinggil sa "Kahalagahan ng Wika at Panitikan" sa loob ng itinakdang oras sa isang kontroladong kapaligiran upang matiyak ang pagiging orihinal ng kanilang likha. Ikatlo, ang mga nakalap na sanaysay ay isinailalim sa pagsusuri gamit ang iba't ibang AI content detection tools, at ang mga resulta ay maingat na itinala at inorganisa para sa susunod na pagsusuri.

Samantala, sa kwalitatibong yugto, isinagawa ang pangangalap ng datos sa pamamagitan ng masusing panayam. Una, bumuo ng semi-structured interview guide na nakatuon sa karanasan, pananaw, at reaksyon ng mga mag-aaral hinggil sa paggamit ng AI content detection tools. Ikalawa, isinagawa ang in-depth interview sa piling sampung (10) respondente sa isang angkop at kompidensyal na kapaligiran. Ang mga panayam ay naitala at isinailalim sa transkripsyon upang maging batayan ng pagsusuri sa susunod na yugto ng pananaliksik.

Pagsusuri ng Datos

Sa kwantitatibong bahagi ng pananaliksik, sinuri ang bisa ng mga AI content detection tools gamit ang mga sukatan na precision, recall, at F1 score upang masukat ang kanilang katumpakan sa pagkilala ng AI-generated at human-written na mga sanaysay. Ginamit din ang False Positive Rate (FPR) at False Negative Rate (FNR) upang matukoy ang antas ng maling klasipikasyon. Upang masuri ang pagkakaiba sa pagganap ng mga AI tools, isinagawa ang mga estadistikal na pagsusuri tulad ng Chi-square test at Analysis of Variance (ANOVA).

Sa kwalitatibong bahagi naman, ginamit ang thematic analysis nina Braun at Clarke (2006) upang suriin ang mga pananaw, karanasan, at saloobin ng mga mag-aaral hinggil sa paggamit ng AI content detection tools. Ang mga panayam ay isinailalim sa transkripsyon, sinundan ng coding, at pagbuo ng mga tema na kumakatawan sa mga karaniwang pattern ng karanasan ng mga kalahok.

Mga Pagsasaalang-alang sa Etika

Isinagawa ang pananaliksik alinsunod sa mga pamantayang etikal upang matiyak ang proteksyon at kapakanan ng mga kalahok. Ang lahat ng respondente ay binigyan ng informed consent form na naglalaman ng layunin ng pag-aaral, boluntaryong paglahok, at karapatang umatras anumang oras nang walang anumang negatibong epekto (Resnik, 2020). Upang mapanatili ang kumpidensyalidad, gumamit ng mga pseudonym o code sa halip na tunay na pangalan ng mga kalahok (Berg, 2012).

Ang lahat ng nakalap na datos ay itinago sa isang ligtas na imbakan na may limitadong akses upang mapangalagaan ang pribadong impormasyon ng mga respondente. Bukod dito, isinagawa ang pag-aaral nang may pagsunod sa mga pamantayan ng akademikong integridad, kabilang ang wastong pagbibigay-kredito sa lahat ng ginamit na sanggunian upang maiwasan ang plagiarism (Roig, 2015).

Bago ang implementasyon ng pag-aaral, nakakuha ng pahintulot mula sa kinauukulang institusyon at naaprubahan ng Cebu Normal University Ethics Review Committee (Approval Code: 1709/2025-03), alinsunod sa mga alituntunin ng etikal na pananaliksik (American Psychological Association, 2017).

**MGA RESULTA AT TALAKAYAN**

Ipinapakita sa bahaging ito ang mga resulta ng nakalap na datos mula kwantitatibo hanggang kwalitatibong yugto ng pag-aaral.

A. Pagsusuri sa Kabisaan ng AI Content Detection Tools sa Pagkilala ng AI-Generated at Human-Written na mga Sanaysay sa Filipino**Talahanayan 1**

Mga Sukatan sa Pagganap ng mga AI Content Detector sa 80% na Threshold

Detector	Precision (%)	Recall (%)	F1 Score (%)
SMODIN	0.0%	0.0%	0.0%
GPT Zero	100.0%	11.39%	20.45%
Justdone	49.59%	75.95%	60.0%
OpenL.io	77.94%	67.09%	72.11%
ISGEN	100.0%	20.25%	33.68%
Zero GPT	92.98%	67.09%	77.94%

Batay sa Talahanayan 1, makikita ang malinaw na pagkakaiba sa pagganap ng mga AI content detection tools sa pagtukoy ng AI-generated at human-written na mga sanaysay sa wikang Filipino. Ang **Zero GPT** ang nakapagtala ng pinakamataas na F1 score (77.94%), na nagpapahiwatig ng balanseng kakayahan sa pagitan ng precision at recall. Sinundan ito ng **OpenL.io (72.11%)**, na nagpapakita rin ng maaasahang performance sa parehong sukatan.

Sa kabilang banda, ang **GPT Zero** at **ISGEN** ay nagpakita ng 100% precision ngunit mababang recall (11.39% at 20.25%), na nangangahulugang bagama't tama ang kanilang mga natukoy, marami pa rin ang hindi nila natutukoy na AI-generated na teksto. Ipinapakita nito na ang mataas na precision lamang ay hindi sapat na batayan ng pagiging epektibo ng isang AI detector kung kulang naman ito sa kakayahang tukuyin ang karamihan ng AI-generated content.

Samantala, ang **SMODIN** ay nagtala ng 0.0% sa lahat ng sukatan, na nagpapakita ng kabiguang makapagbigay ng tamang klasipikasyon sa dataset na ginamit. Ipinahihiwatig nito ang limitasyon ng ilang AI detection tools kapag inilapat sa wikang Filipino, na maaaring dulot ng kakulangan sa sapat na training data para sa mga non-English na teksto. Ang ganitong resulta ay umaayon sa pag-aaral nina Weber-Wulff et al. (2023), na nagsasaad na ang maraming AI detection tools ay hindi pa sapat na maaasahan at may mga limitasyon sa katumpakan sa iba't ibang konteksto ng wika.

Talahanayan 2

Mga Antas sa Confusion Matrix para sa mga AI Content Detector sa 80% na Threshold

AI Detector	True Positive (TP)	False Positive (FP)	Positive True (TN)	Negative False (FN)	Negative Rank
SMODIN	0.00	98.75	96.25	3.75	6
GPT Zero	11.25	87.50	100.00	0.00	5
Justdone	75.00	23.75	23.75	76.25	3
OpenL.io	66.25	32.50	81.25	18.75	2
ISGEN	20.00	78.75	100.00	0.00	4
Zero GPT	66.25	32.5	95.0	5.0	1

Batay sa Talahanayan 2, mas detalyadong naipakita ang pagganap ng mga AI content detectors gamit ang confusion matrix. Ang **Zero GPT** at **OpenL.io** ay nagpakita ng mas balanseng resulta sa pagitan ng true positives at true negatives, na nagpapahiwatig ng mas maaasahang kakayahan sa tamang klasipikasyon ng mga teksto.

Sa kabilang banda, ang **SMODIN** ay may napakataas na false positive rate (98.75%), na nangangahulugang karamihan sa human-written na sanaysay ay maling naituring bilang AI-generated. Ang ganitong uri ng maling klasipikasyon ay may seryosong implikasyon sa akademikong konteksto, sapagkat maaaring magdulot ito ng hindi makatarungang paratang laban sa mga mag-aaral.

Gayundin, ang **GPT Zero** at **ISGEN** ay nagpakita ng mataas na true negative rate ngunit mababang true positive rate, na nagpapahiwatig ng kakulangan sa kakayahang matukoy ang AI-generated na teksto. Ipinapakita nito



na hindi sapat ang pagiging konserbatibo ng isang AI detector kung hindi naman nito natutukoy ang aktwal na AI-generated content.

Sa kabuuan, ipinapakita ng mga resulta na ang pagiging epektibo ng isang AI content detector ay nakasalalay sa balanse ng iba't ibang sukatan, at hindi lamang sa iisang aspeto ng pagganap. Ang mga natuklasang ito ay sinusuportahan ng pag-aaral nina Bender et al. (2021), na nagbibigay-diin sa mga limitasyon ng mga AI system sa pag-unawa sa konteksto ng wika at ang panganib ng labis na pag-asa sa mga ito sa mga sensitibong larangan tulad ng edukasyon.

Upang mapalalim at mapaunlad ang interpretasyon ng mga kwantitatibong natuklasan, isinagawa ang kwalitatibong yugto bilang pangalawang yugto ng *explanatory sequential mixed-methods* na disenyo ng pananaliksik. Alinsunod sa pamamaraang ito nina Creswell at Creswell (2018), ang kwalitatibong datos ay espesipikong ginagamit upang ipaliwanag, palawakin, at bigyan ng kontekstuwal na kahulugan ang mga numerikong resulta na hindi ganap na maipaliwanag ng estadistika. Sa ganitong paraan, ang mga temang lumitaw mula sa in-depth interviews sa mga mag-aaral ng BSEd-Filipino ay nagsisilbing interpretibong lens sa mga natuklasang pagganap ng bawat AI content detection tool. Partikular, ang mga numerikong resulta na nagpapakita ng malawak na pagkakaiba-iba sa precision, recall, at F1 score - kasama na ang napakataas na false positive rate ng SMODIN (98.75%) at ang pangkalahatang mababang pagganap ng karamihan sa mga tool sa wikang Filipino ay ang mga pangunahing resultang pinalawig at ipinaliwanag ng kwalitatibong bahagi ng pag-aaral. Ang integrasyon ng dalawang uri ng datos ay isinasagawa sa antas ng interpretasyon, kung saan ang mga kwalitatibong tema ay nagbibigay ng dahilan, konteksto, at karanasang nagpapatibay sa mga numerikong natuklasan.

B. Pananaw, Saloobin at Karanasan ng mga Mag-aaral sa Epektibo at Patas na Paggamit ng AI Content Detectors sa Akademikong Pagsulat sa Filipino

Inilalahad sa bahaging ito ang mga natuklasang tema at pananaw mula sa pagsusuri ng nakalap na datos sa isinagawang panayam. Gamit ang *thematic analysis* (Braun at Clarke, 2006), natukoy ang mga makabuluhang ugnayan, karanasan, at saloobin ng mga kalahok kaugnay ng paggamit ng mga AI content detector sa pagtukoy ng orihinalidad ng akademikong sulatin.

Tema 1. Limitasyon ng AI Detectors sa Kontekstong Akademiko

Ang temang ito ay direktang nagpapaliwanag sa numerikong resulta ng pag-aaral, lalo na ang kabiguang makapagbigay ng anumang wastong klasipikasyon ng SMODIN (F1 score = 0.0%) at ang pangkalahatang mababang kakayahan ng maraming tool na tukuyin ang AI-generated na teksto.

Kadagdagan pa, ang temang ito ay lumitaw bilang isa sa mga pinakamasidhing usapin mula sa mga kalahok. Madalas nilang banggitin na ang mga AI content detection tools ay hindi pa sapat, hindi ganap na maaasahan, at may mga kakulangan pagdating sa pagiging patas at tiyak - lalo na sa konteksto ng akademikong pagsusulat. Ang mga kalahok ay naghayag ng pagdududa sa katumpakan at konsistensi ng mga tool na ito, at naniniwalang hindi dapat ito maging pangunahing batayan ng akademikong desisyon gaya ng pagtukoy ng plagiyarismo, paglabag sa akademikong integridad, o husay ng gawa. Ang pangunahing isyu ay ang kakulangan sa *transparency*, kakayahan nitong umunawa ng konteksto ng wika, lalo na sa Filipino, at ang posibilidad ng maling akusasyon. Ayon kina K6 at K9:

"Hindi sapat na basehan ang AI detectors kasi may mga pagkakataon na nagkakaroon sila ng false result o false positives. Mas mainam pa rin na dumaan sa masusing pagsusuri ng guro o eksperto." (K6)

"...ang AI ay hindi pa ganap na pulido... nagkakaroon ng inconsistency sa resulta. Kaya't sa usaping akademiko, delikado kung ito lang ang pagbabatayan ng plagiarism o orihinalidad." (K9)

Batay sa mga nakalap na datos, malinaw na isinusulong ng mga kalahok ang papel ng *human judgment* bilang mahalagang bahagi ng pagsusuri ng orihinalidad sa akademikong sulatin. Kapwa nagbigay-babala ang mga kalahok hinggil sa *false positives* at *inconsistency* ng mga resulta mula sa AI tools. Ayon sa mga kalahok, ang mga teknolohiyang ito ay hindi pa hinog upang gamitin bilang pamantayan sa mga sensitibong usapin tulad ng plagiyarismo, na may malalim na epekto sa karera at reputasyon ng mga mag-aaral. Tumutugma ito sa pag-aaral nina Elkhatat et al. (2023) na nagkakaroon ng *inconsistency* o maling klasipikasyon ang mga detector sa *human-*



written na teksto. Dagdag pa nila, huwag umasa lamang sa mga *tool* na ito at gumamit ng karagdagang paraan ng pag-verify upang maiwasan ang hindi makatarungang pagtanggap sa tunay na gawa ng mag-aaral. Bagama't nakatutulong ang *AI* bilang pantulong na kasangkapan, hindi ito sapat upang magsilbing tanging batayan ng hatol. Ang mas ligtas na praktis ay ang pagsama ng *human judgment* sa bawat hakbang ng pagsusuri.

Tema 2. Pamantayan sa Orihinalidad at Pagiging Patas ng AI Detectors

Ang mga pananaw ng mga kalahok hinggil sa kakulangan ng pagiging patas ng mga *tool* ay nauugnay sa mga resulta ng confusion matrix, kung saan ipinakita ng SMODIN ang 98.75% false positive rate — isang numerikong katibayan ng hindi makatarungang paratang laban sa mga mag-aaral na nagsusulat sa wikang Filipino.

Ito ay tumutukoy sa mga batayan o sukatan na ginagamit upang masuri kung gaano kabisa at makatarungan ang mga *AI content detector* sa pagtukoy ng orihinalidad ng isang akademikong sulatin. Ito ay mahalaga sa konteksto ng edukasyon at pananaliksik, kung saan ang integridad ng gawa ay dapat mapanatili. Ayon kina 1 at K4:

"Subalit, hindi ko nakikitang patas itong gamitin upang maging kagamitan sa paghatol sa kredibilidad ng isang katha... masasabi kong hindi ito patas lalo sa akin na alam kong ako mismo ang sumulat ng naturang sanaysay." (K1)

"Sa aking palagay ay hindi po... kung gagamitin natin ang AI detectors bilang basehan... hindi ito maging patas lalo na sa mga mag-aaral na magaling at may kasanayan talaga sa pagbuo ng mga sulatin." (K4)

Ang temang ito ay tumutukoy sa usapin ng pagiging patas at posibleng pagkiling ng mga *AI content detector* sa pagtukoy ng orihinalidad ng sulatin sa wikang Filipino. Batay sa pananaw ng mga kalahok, may mga pagkakataong hindi isinasaalang-alang ng teknolohiya ang konteksto ng wika at istilo ng manunulat. Ipinapahiwatig ng mga obserbasyong ito na ang mekanismo ng *AI detection* ay maaaring may pagkiling, partikular sa mga tekstong nasa wikang Filipino, dahil karamihan sa mga *tool* ay mas sinanay sa wikang Ingles. Ang pahayag na ito ay sinusuportahan sa pag-aaral ni Furze (2024), bagaman may mga tagumpay ang *AI detectors*, maraming *breakdown points* na nagdudulot ng bias at hindi patas na akusasyon, kaya't hindi ito angkop bilang sole integrity check. Sa ganitong diwa, iminumungkahi ang paggamit ng *AI* detektor ay dapat suportado ng pantay-pantay na pananaliksik ng guro, pati ang pagbibigay ng pagkakataong ipaliwanag ng mag-aaral ang kanyang trabaho.

Tema 3. Epekto ng Wikang Filipino sa Katumpakan ng Detection

Ang temang ito ay nagbibigay ng lingguwistikong paliwanag sa pangkalahatang mababang pagganap na naobserbahan sa kwantitatibong yugto, kung saan kahit ang pinakamataas na F1 score (Zero GPT, 77.94%) ay nagpapakita pa rin ng makabuluhang agwat mula sa perpektong klasipikasyon — isang limitasyong naiuugnay ng mga mag-aaral sa pagkiling ng mga *tool* sa wikang Ingles.

Ang temang ito ay tumutukoy sa kung paano naapektuhan ng paggamit ng wikang Filipino ang kakayahan ng *AI content detectors* na matukoy kung ang isang sulatin ay *AI-generated* o *human-written*. Marami sa mga *AI detectors* ay orihinal na binalangkas at sinanay gamit ang mga tekstong nasa wikang Ingles, kaya't maaaring hindi ito ganoon kabisa kapag ginamit sa mga tekstong nasa Filipino. Dahil dito, nagkakaroon ng posibilidad na bumaba ang katumpakan ng pagsusuri, na maaaring magresulta sa maling hatol kung *AI-generated* o orihinal ang isang sulatin. Ayon kina K1 at K5:

"Maaaring may epekto ito sapagkat ang pagkakaalam ko ay mas bihasa sa wikang Ingles ang mga AI detectors... kaya mas madali nilang matukoy kung ito ba ay gawa ng tao o hindi, kumpara sa wikang Filipino." (K1)

"Sa tingin ko mayroon kasi sa English, parang nandoon yung focus ng AI... sa Filipino parang hindi napagtutuunan ng pansin, hindi accurate yung pagdetect ng AI." (K5)

Batay sa mga naging kasagutan, nanindigan ang mga kalahok na may *bias* ang mga *AI* detektor na pabor sa Ingles. Tumutugma ito sa mga pag-aaral na nagpapakita na 60% hanggang 90% ng mga sanaysay ng *non-native English* ay maling naitatalaga bilang gawa ng *AI* at ito ay nagpapahiwatig na dahil sa pagkakaiba sa *sophistication* ng wika, mas madalas na mali ang hatol sa hindi katutubong mga sulatin. Sa konteksto ng Filipino, pinagkakaitan ng mga *detector* ng pantay na pagkilala ng mga sulatin dahil dito, marapat na kilalanin ng guro at pamantasan ang limitasyon at huwag ituring na perpekto ang teknolohiya kapag wikang Filipino ang ginagamit. Ang temang ito ay



sumasalamin sa pang-akademikong usapin na panatilihin ang kamalayan sa *bias* ng mga *AI tool*, lalo na sa *heterogenous* na wika ng mga mag-aaral (Myers, 2023).

Tema 4. Reaksiyon at Sariling Pagtingin ng Manunulat

Ang emosyonal na epektong inilarawan ng mga kalahok ay direktang resulta ng mataas na insidente ng maling klasipikasyon na natuklasan sa kwantitatibong yugto, partikular ang mataas na false positive rates na nagdudulot ng hindi makatarungang paratang sa mga mag-aaral na nagsusulat nang orihinal.

Ito ay tumatalakay sa emosyonal at intelektwal na tugon ng mga mag-aaral bilang mga manunulat, matapos nilang malaman ang resulta ng pagsusuri ng *AI detectors* sa mga sariling gawa. Sa konteksto ng pag-aaral, lumilitaw ang temang ito bilang mahalagang dimensyon ng epekto ng teknolohiya sa karanasan ng mga mag-aaral sa akademikong pagsulat. Maraming mag-aaral ang nakararanas ng pagkalito, pagkadismaya, o kawalan ng tiwala sa sarili kapag ang kanilang orihinal na sulatin ay mali o hindi patas na kinilalang gawa ng *AI*. Sa ganitong sitwasyon, naapektuhan hindi lamang ang kanilang emosyon, kundi maging ang kanilang motibasyon, tiwala sa sarili, at pananaw sa paggamit ng teknolohiya sa edukasyon. Ayon kina K1 at K7:

"Ang unang reaksiyon ko noon ay napatanong ako ng 'Bakit nangyari iyon?' — dahil alam ko namang orihinal kong isinulat ang sanaysay. Kumbaga, ano kaya ang naging basehan ng naturang AI detector bakit ganyan ang lumabas na datos?" (K1)
"Sa una, noong nalaman ko na mula sa AI raw yung ginawa kong sanaysay, ako ay nabigla...Sa una, parang nanlumo, nanghina sapagkat papaano magiging nagmula sa AI kung hindi naman ako gumamit." (K7)

Ipinakikita ng mga siping ito ang malinaw na epekto ng pagkakadetek sa mga sulatin ng mga kalahok bilang gawa ng *AI*. Inilarawan ng mga kalahok ang damdamin ng pagkabigla at panghihinayang dahil sigurado ang mga kalahok na orihinal ang kanilang sinulat. Ayon kay Fowler (2023), ang mga taong maling nasasangkot sa isyu ng *AI plagiarism* ay maaaring makararanas ng mga potensyal na nakapipinsalang epekto sa pagtingin sa kanilang sarili. Ang mga panayam ay nagpatibay na ang ganitong insidente ay nakahahadlang sa tiwala ng manunulat sa kanyang sariling kakayahan. Nakasaad sa mga kalahok ang paninindigan na bagaman hindi nila naramdaman ang pagbabago sa kanilang kakayahan, nakaapekto ito sa kanilang kumpiyansa at pananaw na posibleng sisiyasatin ang kanilang gawa.

Tema 5. Epekto ng *AI Detectors* sa Motibasyon ng mga Mag-aaral na Magsulat

Ang pagbaba ng motibasyon na iniulat ng mga kalahok ay hindi maiuugnay sa iisang insidente — ito ay isang kinalabasan ng sistematikong hindi mapagkakatiwalaang pagganap na natuklasan sa kwantitatibong pagsusuri, kung saan kahit anim na tool ang sinuri ay walang isa man ang nagpakita ng sapat at konsistenteng katumpakan sa konteksto ng wikang Filipino.

Ang paggamit ng *AI content detection tools* sa pagsusuri ng akademikong sulatin ay may layuning mapanatili ang integridad sa edukasyon sa pamamagitan ng pagtukoy sa mga gawaing hindi orihinal. Gayunpaman, batay sa pananaw ng mga kalahok, may hindi inaasahang negatibong epekto ito sa kanilang emosyonal na kalagayan at akademikong motibasyon. Ang takot na matawag na *AI-generated* ang sariling gawa ay humahantong sa pagbaba ng *intrinsic motivation* at *self-efficacy*, dalawang mahahalagang elemento sa matagumpay na pagkatuto. Nabuo sa mga pahayag ng kalahok ang ideya na ang pagsusulat ay hindi na lamang isang paraan ng pagpapahayag ng sarili, kundi isang aktibidad na kailangang makalampas sa mata ng *AI*. Ang ganitong *pressure* ay nagdudulot ng pagkabahala at pangamba, at sa ilang pagkakataon, nawawalan ng gana ang mga mag-aaral na magsulat, kahit pa orihinal at tapat ang kanilang gawa. Ayon kina K3 at K9:

"Nakakawalang gana magsulat kapag alam mong kahit ikaw ang gumawa, may posibilidad na sabihing hindi ito original." (K3)
"Parang nawawalan ako ng tiwala sa sarili ko. Gumawa ako, pero sabi ng tool baka AI raw. So parang, sayang lang ang effort." (K9)

Batay sa mga naging kasagutan ng mga kalahok, ito'y patunay lamang na naapektuhan ang kanilang motibasyon sa pagsusulat. Ayon kay Hirsch (2024), ang paggamit ng mga *AI detection systems* ay maaaring magdulot ng sikolohikal na epekto sa mga mag-aaral, kabilang ang stress at takot na dulot ng maling klasipikasyon ng kanilang orihinal na mga sulatin bilang gawa ng *AI*. Sa pagnanais na masukat ang *bisa* ng mga *AI tools*, kailangang timbangin hindi lamang ang kakayahang makadetekta ng *AI-generated content*, kundi pati ang emosyonal, lingguwistiko, at epekto sa motibasyon sa mga estudyanteng Pilipino na nagsusulat sa wikang Filipino.



Kongklusyon

Batay sa mga natuklasang datos mula sa pinagsamang kwantitatibo at kwalitatibong pagsusuri, malinaw na ang mga kasalukuyang AI content detector ay may malalaking limitasyon sa pagtukoy ng orihinalidad ng akademikong sulatin, partikular sa konteksto ng wikang Filipino. Ipinakita ng kwantitatibong resulta ang hindi pareparehong antas ng katumpakan ng mga tool batay sa precision, recall, at F1 score, habang ang kwalitatibong datos naman ay nagbigay-lalim sa mga numerikal na ito sa pamamagitan ng paglalantad ng aktwal na karanasan ng mga mag-aaral, kabilang ang pagdududa, pagkabahala, at pagbaba ng kumpiyansa dulot ng maling klasipikasyon. Sa pamamagitan ng integrasyon ng dalawang uri ng datos, naipakita ng pag-aaral na ang mga isyung teknikal tulad ng false positives ay may direktang implikasyong emosyonal at akademikong, na hindi lubos na mauunawaan kung iisang metodo lamang ang ginamit.

Ipinahihiwatig ng pag-aaral na may umiiral na bias ang mga AI detection tools laban sa mga tekstong hindi nakasulat sa wikang Ingles, na nagreresulta sa hindi patas na pagsusuri sa mga sulating Filipino. Dahil dito, ang paggamit ng mga AI detector bilang pangunahing batayan ng akademikong paghatol ay hindi pa angkop sa kasalukuyang konteksto. Sa kabuuan, ang pag-aaral na ito ay hindi lamang nag-ambag ng ebidensyang estadistikal hinggil sa limitasyon ng mga AI detection tools, kundi nagbigay rin ng kontekstuwal at makataong pag-unawa sa kanilang epekto, na nagpapatibay sa kahalagahan ng maingat, kritikal, at kontekstuwal na paggamit ng teknolohiya kasama ng human judgment upang mapanatili ang katarungan, integridad, at tiwala sa loob ng edukasyonal na sistema.

Rekomendasyon

Iminumungkahi ng pag-aaral na ang paggamit ng mga AI content detection tools ay maaaring isaalang-alang bilang pantulong na mekanismo sa pagsusuri ng orihinalidad ng akademikong sulatin, sa halip na magsilbing nag-iisang batayan ng akademikong desisyon. Makabubuting isaalang-alang ang paggamit ng masusing human judgment, lalo na mula sa mga guro at eksperto sa wika, upang matiyak ang mas makatarungan at kontekstuwal na pagsusuri ng mga sulatin sa wikang Filipino.

Dagdag pa rito, maaaring magsilbing batayan ang mga natuklasan ng pag-aaral para sa mga institusyong pang-edukasyon sa pagbuo ng mga patakaran hinggil sa responsableng paggamit ng AI detection tools. Para naman sa mga mananaliksik, maaaring magsagawa pa ng karagdagang pag-aaral na nakatuon sa iba pang wika, konteksto, at uri ng sulatin upang higit na mapalawak ang pag-unawa sa bisa at limitasyon ng mga naturang teknolohiya.

MGA SANGGUNIAN

- Abdali, S., Anarfi, R., Barberan, C., & He, J. (2024). *Decoding the AI pen: Techniques and challenges in detecting AI-generated text*. arXiv. <https://arxiv.org/abs/2403.05750>
- American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. <https://www.apa.org/ethics/code>
- Angeles, J. B., Matuguina, S. P., Camarillo, S. C., & Dimacali, J. C. (2024). Students' perception of the use of AI detector systems by faculty members in determining the originality of submitted academic requirements. *Proceedings of the ACM Conference*. <https://doi.org/10.1145/3675585.3675587>
- Bender, E. M., McMillan-Major, A., Shmitchell, S., & Geburu, T. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Berg, B. L. (2012). *Qualitative research methods for the social sciences* (8th ed.). Pearson.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Coldwell, W. (2024, December 15). "I received a first but it felt tainted and undeserved": Inside the university AI cheating crisis. *The Guardian*. <https://www.theguardian.com/technology/2024/dec/15/i-received-a-first-but-it-felt-tainted-and-undeserved-inside-the-university-ai-cheating-crisis>
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.
- Creswell, J. W., & Creswell, J. D. (2018). Mixed methods procedures. In *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed., pp. 213–246). SAGE Publications.



- De La Salle University. (2023). *Policy on the use of generative artificial intelligence in education*. <https://www.dlsu.edu.ph/wp-content/uploads/pdf/provost/forms/policy-for-generative-ai-in-education.pdf>
- Dingal, S. M. L., Balbin, C. J. P., Buenafe, S. G. L., Buenavista, P. E. M., Cabrera, C. A., Lorica, V. M. T., Olalo, A. M., Pataganao, L. A. A., Wahing, M. R. A., Clamares, K. J. M., & Pelandas, A. M. O. (2024). Artificial intelligence usage and its influence on students' academic writing. *International Journal of Research and Innovation in Social Science*, 8(4), 1621–1627. <https://doi.org/10.47772/IJRIS.2024.804216>
- Elkhatat, A. M., Elsaid, K., & Al-Meer, S. (2023). Evaluating the efficacy of AI content detection tools. *International Journal for Educational Integrity*, 19(1). <https://doi.org/10.1007/s40979-023-00140-5>
- Estrellado, C., & Miranda, J. (2023). Artificial intelligence in the Philippine educational context. *International Journal of Scientific and Research Publications*, 13(5), 16–18. <https://doi.org/10.29322/IJSRP.13.05.2023.p13704>
- Fabro, R. B. B., Rivera, J. C., Sambrano, L. C., Dinoy, L. J. M., Alcozer, N. G., & Agustin, M. C. (2024). Perceptions and extent of utilization of generative artificial intelligence (AI) among Filipino students. *International Journal of Education and Research*, 12(7), 107–126. <https://ijern.com/journal/2024/July-2024/09.pdf>
- Fowler, G. (2023, August 14). What to do when you're accused of AI cheating. *The Washington Post*. <https://www.washingtonpost.com/technology/2023/08/14/prove-false-positive-ai-detection-turnitin-gptzero/>
- Furze, L. (2024, April 9). AI detection in education is a dead end. <https://leonfurze.com/2024/04/09/ai-detection-in-education-is-a-dead-end/>
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv*. <https://doi.org/10.1101/2022.12.23.521610>
- Giray, L. (2024). The problem with false positives: AI detection unfairly accuses scholars of AI plagiarism. *The Serials Librarian*, 86(1–2), 1–10. <https://doi.org/10.1080/0361526X.2024.2433256>
- Giray, L., Sevnarayan, K., & Madiseh, F. R. (2025). Beyond policing: AI writing detection tools, trust, and academic integrity. *Internet Reference Services Quarterly*. <https://doi.org/10.1080/10875301.2024.2437174>
- Hirsch, A. (2024). AI detectors: An ethical minefield. *Center for Innovative Teaching and Learning*. <https://citl.news.niu.edu/2024/12/12/ai-detectors-an-ethical-minefield/>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. M. T., Cukorova, M., Bittencourt, I. I., & Koedinger, K. (2021). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32, 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- Liang, W., Yuksekogonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *arXiv*. <https://doi.org/10.48550/arXiv.2304.02819>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). American Council on Education & Macmillan.
- Montalan, J. R., Layacan, J. P., Africa, D. D., Flores, R. I., Lopez, M. T., II, Magsajo, T. D., Cayabyab, A., & Tjhi, W. C. (2025). *Batayan: A Filipino NLP benchmark for evaluating large language models*. arXiv. <https://arxiv.org/abs/2502.14911>
- Myers, A. (2023, May 15). AI detectors biased against non-native English writers. *Stanford HAI*. <https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers>
- Nam, J. (2023, November 22). 56% of college students have used AI on assignments or exams. *BestColleges*. <https://www.bestcolleges.com/research/most-college-students-have-used-ai-survey/>
- Ordonez, J. R., Sadiasa, J. B. S., & Soriaso, B. D. F. (2024). The rise of AI in education. *GreenPrints*. https://greenprints.dlshsi.edu.ph/grade_12/649/
- Pecasion, S., Papelleras, S. J., Pedrozo, P., & Allawan, J. G. L. (2023). Silver lining of ChatGPT artificial intelligence tool: Through the lens of department of teachers education in Panabo City, Philippines. *Journal of Learning and Educational Policy*, 4(1), 19–29. <https://doi.org/10.55529/jlep.41.19.29>
- Resnik, D. B. (2020). *The ethics of research with human subjects*. Springer.
- Roig, M. (2015). *Avoiding plagiarism, self-plagiarism, and other questionable writing practices*. U.S. Department of Health & Human Services.
- Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). *Can AI-generated text be reliably detected?* arXiv. <https://arxiv.org/abs/2303.11156>
- Trist, E., & Bamforth, K. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4(1), 3–38. <https://doi.org/10.1177/001872675100400101>



- University of the Philippines Open University. (2023). *Guidelines on the ethical use of generative AI*. <https://www.upou.edu.ph/news/upou-releases-guidelines-on-ai-use-for-teaching-and-learning/>
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J. G., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19, Article 26. <https://doi.org/10.1007/s40979-023-00146-z>
- Wu, H., & Flanagan, T. (2023). The limits of AI content detectors. *Journal of Student Research*. <https://www.jsr.org/hs/index.php/path/article/view/5064>