บทบรรณาธิการ

เมื่อ AI ตอบทุกคำถามแต่ไม่รับผิดชอบ: ความเสี่ยงและความจำเป็นเร่งด่วนของกรอบจริยธรรมและธรรมาภิบาล AI

ประเด็นร้อนแรงในโลกของการใช้เทคโนโลยีดิจิทัลในธุรกิจ ณ ปี ค.ศ. 2026 ก็ยังคงหนีไม่พ้นเรื่องของการใช้ปัญญาประดิษฐ์ (Artificial Intelligence) หรือที่เรียกกันสั้น ๆ ว่า AI ตั้งแต่ที่ OpenAI ได้เปิดตัว ChatGPT ในปลายปี ค.ศ. 2022 AI ก็กลายเป็นเทคโนโลยีที่เข้าถึงได้ง่าย ใช้งานได้ง่าย และผู้คนทั่วโลกโดยเฉพาะกลุ่มพนักงานออฟฟิศต่างหันหน้าเข้าหา Generative AI รายงานเชิงสำรวจเรื่อง State of Organizations 2026 ของ McKinsey พบว่า องค์กรถึงร้อยละ 88 มีการนำ Generative AI มาใช้ในส่วนใดส่วนหนึ่งของการทำงาน (McKinsey & Company, 2026) ในขณะที่รายงานจาก Wharton Human-AI Research ระบุว่า ร้อยละ 82 ของผู้นำองค์กรใช้ Generative AI เป็นประจำทุกสัปดาห์ และ ร้อยละ 88 ของผู้นำองค์กรคาดว่าเพิ่มการลงทุนเพื่อการใช้งาน Generative AI ในปีค.ศ. 2026 (Wharton Human-AI Research & GBK Collective, 2025)

บทความและงานวิจัยมากมายต่างมุ่งความสนใจไปที่การนำ AI ไปประยุกต์ใช้ในธุรกิจเพื่อสร้างความได้เปรียบในการแข่งขัน เพื่อการลดต้นทุน เพื่อการเพิ่มประสิทธิภาพในการทำงาน และเพื่อประโยชน์อื่น ๆ อีกมากมาย ในบทความนี้ผู้เขียนจึงอยากชวนมอง AI จากอีกมุมหนึ่ง คือมุมของข้อเสีย ผลกระทบเชิงลบ หรือความเสี่ยงที่อาจจะตามมา โดยผู้เขียนขอแบ่งเป็นประเด็นต่าง ๆ ดังนี้

1) ความเสี่ยงในเรื่องของข้อมูลรั่วไหล ประเด็นนี้ได้รับการพูดถึงในวงกว้าง และนับเป็นเรื่องน่ายินดีที่ผู้คนเริ่มมีความตระหนักถึงเรื่องความปลอดภัยของข้อมูล แต่อย่างไรก็ตาม เพื่อแลกกับความสะดวกและรวดเร็ว พนักงานออฟฟิศจำนวนมากนำข้อมูลจริงที่เป็นข้อมูลจากหน่วยงานใส่เข้าไปให้กับ GenAI ทำการวิเคราะห์และประมวลผล ซึ่งการนำข้อมูลเข้าไปให้กับ GenAI ในลักษณะนี้เกิดขึ้นจริงมาแล้ว โดยกรณีที่เป็นข่าวดังไปทั่วโลกได้แก่กรณีของบริษัท Samsung Semiconductor ที่มีการอนุญาตให้พนักงานใช้ ChatGPT ในการช่วยทำงานได้ แต่ปัญหาที่ตามมาอย่างรวดเร็วและส่งผลเสียอย่างมากคือ พนักงานนำซอร์สโค้ดที่พัฒนากันภายในส่งเข้าไปให้ ChatGPT ตรวจสอบหาข้อผิดพลาด และมีพนักงานอีกรายนำบันทึกการประชุมภายในที่มีเนื้อหาเป็นความลับของบริษัทเข้าไปให้ ChatGPT ช่วยสรุปเป็นรายงาน *เหตุการณ์ทั้งหมดเกิดขึ้นภายในเวลาไม่ถึงหนึ่งเดือนหลังบริษัทอนุญาตให้ใช้งาน และนำไปสู่การสั่งห้ามใช้ ChatGPT ทั่วทั้งองค์กรในเวลาต่อมา* (Jeong, 2023) ในกรณีนี้ถึงแม้ตัว GenAI จะไม่นำข้อมูลเหล่านี้ไปแจ้งแก่ผู้ใช้งานของบริษัทคู่แข่งโดยตรง แต่หากผู้ใช้งานของบริษัทคู่แข่งมีการตั้งคำถามถึงการวางแผนกลยุทธ์หรือการเขียนโค้ดที่มีลักษณะใกล้เคียงกัน มีบริบทที่คล้ายกัน GenAI ก็อาจจะนำข้อมูลที่ได้เรียนรู้มาก่อนหน้ามาประกอบในการสร้างคำแนะนำให้กับบริษัทคู่แข่งได้

2) ปัญหาเกี่ยวกับการสร้างข้อมูลปลอม ข้อมูลเท็จโดย AI ซึ่งในปัจจุบันมีการบัญญัติศัพท์เฉพาะสำหรับเรียกปรากฏการณ์นี้ว่า AI Hallucination *จากประสบการณ์ของผู้เขียนในฐานะอาจารย์มหาวิทยาลัย (ข้อสังเกตเชิงประจักษ์ส่วนบุคคล มิใช่ผลการวิจัยเชิงระบบ)* ผู้เขียนเองได้เห็นนักศึกษาจำนวนมากใช้ GenAI ในการช่วยทำรายงานโดยไม่ได้ตรวจสอบข้อมูล และพบว่าแหล่งอ้างอิงหลายรายการไม่มีอยู่จริง เป็นเพียงรายการที่ AI สร้างขึ้นมาโดยนำชื่อเรื่อง ชื่อผู้แต่ง และชื่อวารสารหรือเว็บไซต์ต่าง ๆ มาผสมกันอย่างแนบเนียน ในบริบทของธุรกิจ เหตุการณ์หนึ่งที่เป็นข่าวใหญ่ระดับนานาชาติ ได้แก่กรณีของบริษัท Deloitte ออสเตรเลีย ที่ได้รับการว่าจ้างจากรัฐบาลออสเตรเลียให้จัดทำรายงานทบทวนระบบสวัสดิการมูลค่ากว่า 290,000 ดอลลาร์สหรัฐ โดยภายหลังส่งมอบรายงานแล้ว นักวิจัยจากมหาวิทยาลัยซิดนีย์ตรวจพบว่ารายงานฉบับดังกล่าวมีแหล่งอ้างอิงทางวิชาการที่ไม่มีอยู่จริงหลายรายการ รวมถึงคำพิพากษาที่ถูกอ้างอิงผิดจากคดีที่ไม่เคยมีคำตัดสินเช่นนั้นจริง ทำให้ Deloitte ต้องคืนเงินบางส่วนแก่รัฐบาลและเปิดเผยภายหลังว่ามีการใช้ระบบ Generative AI ช่วยในการจัดทำรายงาน (McGuirk, 2025) ปัญหาดังกล่าวเป็นปัญหาที่เกิดขึ้นได้ในทุกครั้งของการใช้งาน GenAI เนื่องจากในทางเทคนิคแล้ว GenAI เป็นเพียงซอฟต์แวร์

ที่ถูกพัฒนาขึ้นมาโดยมีฟังก์ชันความน่าจะเป็นทางสถิติเป็นแกนหลัก ซึ่งถูกนำไปประยุกต์ใช้กับการเรียนรู้ภาษาและการสร้างสรรค์เนื้อหาขึ้นมาใหม่ หากคุณนำปัญหานี้ไปถาม GenAI มันจะบอกคุณว่า AI ไม่ได้ผิดและไม่ได้ล้มเหลวในการทำงานตามคำสั่งเลย แต่มนุษย์หรือผู้ใช้งานต่างหากที่ล้มเหลวในการใช้วิจารณญาณและตรวจสอบเนื้อหาที่ AI สร้างขึ้น

3) ปัญหาเกี่ยวกับการนำ AI ไปใช้สร้างสิ่งที่ผิด เช่น คลิปวิดีโอปลอมเพื่อสร้างกระแสการโจมตีทางการเมือง หรือคลิปวิดีโอปลอมที่มีวัตถุประสงค์มุ่งทำลายชื่อเสียงผู้อื่น ซึ่งถือเป็นปัญหาที่นับได้ว่าเป็นวิกฤตและเป็นภัยสังคม หากโลกอินเทอร์เน็ตเต็มไปด้วยคอนเทนต์ปลอมเหล่านี้ที่นับวันจะถูกสร้างได้แนบเนียนและเหมือนจริงมากขึ้นเรื่อย ๆ จนมนุษย์ไม่สามารถแยกแยะได้ด้วยตาเปล่าว่าเป็นภาพเคลื่อนไหวที่ AI สร้างขึ้นหรือเป็นภาพของจริง งานวิจัยสายสังคมศาสตร์เรียกปรากฏการณ์นี้ว่า Infocalypse ซึ่งแตกต่างจาก AI Hallucination ตรงที่ปัญหานี้เกิดจากความตั้งใจของมนุษย์ โดยที่ AI เพียงทำตามคำสั่งของผู้ใช้งานเท่านั้น โดยไม่สามารถแยกแยะได้ว่าผู้ใช้งานที่สั่งให้สร้างคลิปวิดีโอนั้นมีเจตนาที่จะนำไปสร้างความเสียหาย เสื่อมเสียชื่อเสียง หรือสร้างกระแสสังคมในทางที่ผิดหรือไม่ ดังนั้นปัญหานี้จึงไม่ใช่ความผิดของตัว AI โดยตรง แต่เป็นปัญหาสำคัญที่ชี้ให้เห็นว่าเราจำเป็นต้องมีกรอบจริยธรรมเพื่อกำกับการใช้งาน AI หากเราพิจารณาดูแล้วจะพบว่าปัญหาในข้อ 2 และข้อ 3 นั้นมาจากพฤติกรรมมนุษย์เป็นหลัก ดังนั้นเราจึงต้องการกรอบกำกับดูแลที่ครอบคลุมทั้งสองมิติไปพร้อมกัน คือ การพัฒนาเครื่องมือทางเทคนิคเพื่อจำกัดการใช้งานในทางที่ผิด ควบคู่ไปกับการสร้างความรับผิดชอบและทักษะการรู้เท่าทันสื่อในตัวผู้ใช้งานเอง มิเช่นนั้น AI จะถูกใช้เป็นเครื่องมือที่สร้างปัญหาสังคมได้เป็นอย่างมาก

4) ปัญหาต่อพัฒนาการของสมอง ซึ่งโดยส่วนตัวของผู้เขียนมองว่าเป็นปัญหาสำคัญของมวลมนุษยชาติ เนื่องจากเด็กที่เกิดมาในเจเนอเรชันแอลฟาและเบตาเป็นกลุ่มที่เติบโตมาพร้อมกับการใช้ GenAI ตั้งแต่วัยเยาว์ *นักวิจัยจำนวนหนึ่งเริ่มแสดงความกังวล (แม้หลักฐานเชิงประจักษ์ระยะยาวยังอยู่ระหว่างการศึกษา) ว่า* หากเด็กกลุ่มนี้พึ่งพา GenAI มากเกินไป อาจส่งผลเสียต่อพัฒนาการของสมอง เช่น ความสามารถในการเรียบเรียงและการเขียนลดลง กล้ามเนื้อมือไม่ได้รับการฝึกฝนเท่าที่ควร ความจำใช้งาน (working memory) ถูกใช้น้อยลงเนื่องจากพฤติกรรมหลักเป็นเพียงการคัดลอกและวางข้อความ ในทางวิทยาศาสตร์สมอง ปรากฏการณ์นี้เรียกว่า Cognitive Outsourcing ซึ่งหมายถึงการที่มนุษย์มอบหมายงานด้านการคิด ตรรกะ และการสังเคราะห์เหตุผลให้เครื่องมือภายนอกทำแทน ซึ่งอาจส่งผลให้สมองในส่วนที่เกี่ยวข้องกับกระบวนการคิด (Cognitive function) ทำงานน้อยลงและอ่อนแอลงในระยะยาว

ประเด็นปัญหาและความเสี่ยงที่กล่าวไปข้างต้นเป็นเพียงส่วนหนึ่งเท่านั้น ยังมีประเด็นอื่นที่ไม่อาจกล่าวได้หมดในพื้นที่จำกัดนี้ ที่สำคัญกว่านั้นคือ ทั้งสี่ประเด็นนี้มิใช่เพียงข้อสังเกตเชิงปรากฏการณ์ แต่ล้วนมีรากฐานอยู่ในงานวิจัยสายระบบสารสนเทศ (Information Systems) ที่ตีพิมพ์ในวารสารชั้นนำมาก่อนยุค Generative AI จะแพร่หลายเสียอีก งานศึกษาเรื่อง shadow IT ในวารสาร MIS Quarterly Executive และ Business Information Review ได้วางรากฐานแนวคิดเรื่องการรั่วไหลของความรู้ผ่านเทคโนโลยีที่ไม่ได้รับอนุญาตไว้ตั้งแต่ก่อนยุค GenAI งานวิจัยเรื่อง algorithm aversion และ algorithm appreciation ที่ตีพิมพ์ใน MIS Quarterly (Turel & Kalhan, 2023) และเรื่องการมอบหมายงานเชิงปัญญาให้ AI (Cognitive delegation) ใน Information Systems Research (Fügener et al., 2022) ได้ฉายภาพกลไกทางจิตวิทยาที่อยู่เบื้องหลังการเชื่อหรือไม่เชื่อคำแนะนำของ AI อย่างไม่สมเหตุสมผล ซึ่งอธิบายได้ว่าเหตุใดผู้ใช้งานจึงล้มเหลวในการตรวจสอบ AI Hallucination ดังกรณี Deloitte ในขณะที่งานวิจัยเรื่องข่าวปลอมใน MIS Quarterly (Moravec et al., 2019) และ Information Systems Research (Moravec, Kim & Dennis, 2020) ก็ได้ทดสอบมาตรการเชิงแพลตฟอร์มเพื่อต้านทานข้อมูลเท็จมาก่อนที่ปัญหา Infocalypse จาก Generative AI จะทวีความรุนแรงขึ้น ส่วนประเด็นผลกระทบทางปัญญาความจำนั้น มีรากฐานจากแนวคิด cognitive offloading และปรากฏการณ์ Google effect ที่ถูกศึกษาอย่างเป็นระบบในวารสาร Science (Sparrow et al., 2011) และ Trends in Cognitive Sciences (Risko & Gilbert, 2016)

สิ่งที่งานวิจัยเหล่านี้ชี้ให้เห็นร่วมกันคือ ปัญหาทั้งสี่ประการที่กล่าวมาไม่ใช่เรื่องใหม่ในเชิงทฤษฎี หากแต่ Generative AI ได้ขยายขนาดและความเร็วของปัญหาเดิมที่มีอยู่แล้วให้รุนแรงขึ้นอย่างก้าวกระโดด คำถามเชิงวิจัยที่สาขาระบบสารสนเทศจำเป็นต้องตอบจึงไม่ใช่เพียง "ควรมีกรอบจริยธรรมหรือไม่" แต่คือกลไกการกำกับดูแลแบบใดที่ได้ผลจริงในบริบทของ Generative AI ผู้เขียนหวังว่าประเด็นเหล่านี้จะเป็นตัวอย่างที่ชี้ให้เห็นถึงความสำคัญเร่งด่วนของการกำหนดกรอบจริยธรรม จรรยาบรรณ และธรรมาภิบาลกำกับการใช้งาน AI ในอนาคต

งานวิจัยล่าสุดในวารสาร European Journal of Information Systems ได้เริ่มเรียกร้องให้สาขาระบบสารสนเทศให้ความสำคัญกับ "ด้านมืดของ AI" (the dark side of AI) อย่างเป็นระบบมากขึ้น (Mikalef, Conboy, Lundström & Popovič, 2022) และงานวิจัยใน Journal of Strategic Information Systems ก็ได้เสนอกรอบการวิจัยด้านธรรมาภิบาล AI ที่ยังต้องการการต่อยอดอีกมาก (Papagiannidis, Mikalef & Conboy, 2025) บทความนี้จึงเป็นอีกหนึ่งเสียงที่สนับสนุนให้มีงานวิจัยเชิงประจักษ์เพิ่มเติมในประเด็นต่อไปนี้โดยเฉพาะ:

1. กลไกควบคุมการรั่วไหลของข้อมูลผ่าน Shadow AI — องค์กรควรออกแบบนโยบายและเครื่องมือทางเทคนิคแบบใดเพื่อให้พนักงานยังคงได้รับประโยชน์จาก GenAI โดยไม่ต้องแลกมาด้วยความเสี่ยงด้านข้อมูลความลับ และนโยบายเหล่านั้นมีผลต่อพฤติกรรมการใช้งานจริงอย่างไร

2. โพรโทคอลการตรวจสอบผลลัพธ์จาก GenAI ในงานวิชาชีพ — วิชาชีพที่ต้องอาศัยความถูกต้องสูง เช่น งานที่ปรึกษา งานบัญชี งานกฎหมาย ควรมีมาตรฐานการตรวจทาน (verification protocol) สำหรับเนื้อหาที่ผลิตหรือร่วมผลิตโดย AI ในรูปแบบใด และมาตรฐานดังกล่าวควรอยู่ภายใต้การกำกับขององค์กรวิชาชีพหรือกฎหมาย

3. ประสิทธิผลของเครื่องมือต้านข้อมูลเท็จ (anti-fake tools) ระดับแพลตฟอร์ม — ต่อยอดจากงานวิจัยเรื่อง System 1/System 2 interventions (Moravec et al., 2020) เครื่องมือตรวจจับ deepfake และ AI-generated content ในระดับแพลตฟอร์มมีประสิทธิผลเพียงใดเมื่อเทียบกับเทคนิคการสร้างคอนเทนต์ปลอมที่พัฒนาขึ้นอย่างต่อเนื่อง

4. ผลกระทบระยะยาวของการพึ่งพา GenAI ต่อพัฒนาการทางปัญญาในเด็กและเยาวชน — จำเป็นต้องมีงานวิจัยเชิงระยะยาว (longitudinal) ในกลุ่มเจเนอเรชันแอลฟาและเบตาโดยเฉพาะ เนื่องจากหลักฐานปัจจุบันส่วนใหญ่ยังจำกัดอยู่ในบริบทของผู้ใหญ่และเทคโนโลยีดิจิทัลทั่วไป ยังไม่ครอบคลุมถึง Generative AI โดยตรง

5. รูปแบบธรรมาภิบาลที่เปรียบเทียบได้ระหว่างการกำกับดูแลตนเองของแพลตฟอร์ม การกำกับโดยสภาวิชาชีพ และการกำกับโดยกฎหมาย — รูปแบบใดมีต้นทุนการบังคับใช้ต่ำที่สุดและมีประสิทธิผลสูงสุดในบริบทของแต่ละอุตสาหกรรม

กล่าวโดยสรุป AI ได้กลายเป็นส่วนหนึ่งของชีวิตประจำวันของผู้คนไปแล้ว โดยเฉพาะในบริบทของธุรกิจ ไม่ว่าจะเป็นการบริหารจัดการธุรกิจ การตลาด การจัดการ IT infrastructure ไปจนถึงกระบวนการจ้างงาน ดังนั้นคงปฏิเสธไม่ได้ว่าการใช้ AI โดยไม่คำนึงถึงขอบเขตและผลลัพธ์ที่จะตามมา อาจสร้างความเสียหายและส่งผลกระทบต่อสังคมในวงกว้างได้ บทความนี้จึงต้องการเน้นย้ำว่า งานวิจัยในเชิงการยอมรับใช้งานและการประยุกต์ใช้เพียงอย่างเดียวอาจไม่เพียงพออีกต่อไป สาขาระบบสารสนเทศซึ่งมีรากฐานทางทฤษฎีและระเบียบวิธีที่เหมาะสมอยู่แล้วในการศึกษาปฏิสัมพันธ์ระหว่างมนุษย์กับเทคโนโลยี จึงอยู่ในตำแหน่งที่เหมาะสมที่สุดที่จะเป็นผู้นำการวิจัยด้านธรรมาภิบาล AI ในทศวรรษนี้ — ทั้งในแง่ของความเหมาะสม ผลกระทบ และความยั่งยืนของการกำกับดูแลการใช้งาน AI ต่อไป

ลัดดาวัลย์ แก้วกิติพงษ์, บรรณาธิการ

doi: 10.14456/jisb.2026.1

บรรณานุกรม

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2019). Cognitive challenges in human-AI collaboration: Investigating the path towards productive delegation. *Forthcoming, Information Systems Research.* https://dx.doi.org/10.2139/ssrn.3368813

Haag, S., & Eckhardt, A. (2017). Shadow IT. *Business & Information Systems Engineering, 59*(6), 469-473. https://doi.org/10.1007/s12599-017-0497-x

Jeong, D. (2023, March 30). 우려가 현실로…삼성전자, 챗GPT 빗장 풀자마자 '오남용' 속출 [Concerns become reality: Misuse cases erupt as soon as Samsung Electronics lifts ChatGPT ban]. The Economist Korea. https://economist.co.kr/article/view/ecn202303300057?s=31

Mallmann, G. L., Maçada, A. C. G., & Oliveira, M. (2018). The influence of shadow IT usage on knowledge sharing: An exploratory study with IT users. *Business Information Review, 35*(1), 17-28. https://doi.org/10.1177/0266382118760143

McGuirk, R. (2025, October 7). *Deloitte to partially refund Australian government for report with apparent AI-generated errors.* Associated Press. https://apnews.com/article/australia-ai-errors-deloitte-ab54858680ffc4ae6555b31c8fb987f3

McKinsey & Company. (2026). *The state of organizations 2026.* https://www.mckinsey.com/~/media/mckinsey/business%20functions/people%20and%20organizational%20performance/our%20insights/the%20state%20of%20organizations/2026/the-state-of-organizations-2026.pdf

Mikalef, P., Conboy, K., Lundström, J. E., & PopoviČ, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. *European Journal of Information Systems, 31*(3), 257-268. https://doi.org/10.1080/0960085X.2022.2026621

Moravec, P. L., Kim, A., & Dennis, A. R. (2020). Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media. *Information Systems Research, 31*(3), 987-1006. https://doi.org/10.1287/isre.2020.0927

Moravec, P. L., Minas, R. K., & Dennis, A. R. (2019). Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense At All1. MIS quarterly, 43(4), 1343-1360. https://doi.org/10.25300/MISQ/2019/15505

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems, 34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in cognitive sciences, 20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *science, 333*(6043), 776-778. https://doi.org/10.1126/science.1207745

Turel, O., & Kalhan, S. (2023). Prejudiced against the machine? Implicit associations and the transience of algorithm aversion. *Mis Quarterly, 47*(4), 1369-1394. https://doi.org/10.25300/MISQ/2022/17961

Wharton Human-AI Research & GBK Collective. (2025, October 28). *2025 AI adoption report: Gen AI fast-tracks into the enterprise. Knowledge at Wharton*. https://knowledge.wharton.upenn.edu/special-report/2025-ai-adoption-report/

Editorial

When AI Answers Every Question but Takes No Responsibility: Risks and the Urgent Need for an Ethical Framework and AI Governance

The hottest topic in digital business technology in 2026 remains, unsurprisingly, artificial intelligence (AI). Since OpenAI launched ChatGPT in late 2022, AI has become widely and easily accessible — and people around the world, office workers in particular, have rapidly turned to generative AI. According to McKinsey's State of Organizations 2026 survey, 88% of organizations are now deploying generative AI in at least part of their operations (McKinsey & Company, 2026). Meanwhile, 82% of enterprise leaders report using generative AI on a weekly basis as of 2025, with 88% anticipating an increase in Gen AI investment over the next 12 months (Wharton Human-AI Research & GBK Collective, 2025).

A great deal of existing research and commentary focuses on how businesses apply AI to gain competitive advantage, cut costs, and improve efficiency. This article instead invites readers to consider AI from the opposite direction: its downsides, negative consequences, and associated risks. The discussion is organized around four issues.

1) The risk of data leakage. This issue has already received broad public attention, and it is encouraging that awareness of data security has grown. Yet in exchange for convenience and speed, a significant share of office workers continue to feed real organizational data into generative AI tools for analysis and processing. This is not hypothetical — the best-known global case is Samsung Semiconductor, which permitted employees to use ChatGPT for work assistance. Within weeks, employees had pasted internally developed source code into ChatGPT to debug errors, and a separate employee submitted confidential internal meeting notes for the tool to help draft a summary. All of this occurred within roughly a month of the policy taking effect, prompting Samsung to subsequently ban ChatGPT company-wide (Jeong, 2023). Even though the AI tool itself does not directly disclose this data to users from the firm’s competitors, if the users later pose strategic or coding questions in a similar context, the model could draw on patterns learned from such inputs when generating its responses — creating an indirect channel for competitive leakage.

2) The problem of AI-generated false information. This phenomenon is known under the term: AI hallucination. As a university lecturer, the author has personally observed — though this is anecdotal classroom observation rather than systematic research — many students using generative AI to help write reports without verifying the output, only to find that many of the cited references do not actually exist; they are fabrications assembled by the AI from plausible-sounding titles, author names, and publication venues. In the business world, the most prominent recent case is Deloitte Australia, which was commissioned by the Australian government to produce a roughly $290,000 review report on its welfare compliance system. After delivery, a University of Sydney researcher discovered the report contained fabricated academic references and a misattributed federal court quote, prompting Deloitte to issue a partial refund and later disclose that generative AI had been used in preparing the report (McGuirk, 2025). This kind of failure can occur in any use of generative AI, because technically, such systems are statistical-probability engines applied to language modeling and content generation — nothing more. If you ask the AI itself about this problem, it will tell you, accurately, that

it did not fail to follow instructions; rather, the human user failed to exercise judgment and verify the content the AI produced.

3) The misuse of AI to create harmful content — for example, fake videos designed to fuel political attacks or destroy someone's reputation. This is a genuine societal crisis in the making: as fabricated content becomes more convincing and harder to distinguish from reality, the internet risks becoming saturated with material indistinguishable from authentic footage. Social science researchers call this phenomenon the "Infocalypse." It differs from AI hallucination in one key respect — this problem stems from human intent. The AI simply follows the user's instructions, with no way to discern whether the person requesting the video intends to cause harm, damage someone's reputation, or manipulate public opinion. The fault, therefore, does not lie with the AI itself — but this is precisely why it points to an urgent need for an ethical framework. Since the root causes in both Issue 2 and Issue 3 are fundamentally human, effective governance is needed to address both dimensions simultaneously: technical tools that limit misuse, paired with accountability structures and media literacy that shape how users behave. Without such a framework, AI risks becoming a powerful instrument for social harm.

4) The risk to human's cognitive development. The author personally regards this as one of humanity's most consequential challenges. Children of Generation Alpha and Generation Beta, in particular, are growing up with generative AI as a constant presence from early childhood. A growing number of researchers have raised concerns — though long-term empirical evidence is still emerging — that excessive reliance on generative AI among children could impair cognitive development: weaker composition and writing ability, underdeveloped fine motor skills, and reduced use of working memory, since the dominant behavior pattern becomes simply copying and pasting text. Cognitive science refers to this as "cognitive outsourcing" — offloading thinking, logical reasoning, and synthesis to external tools, which may, over time, weaken the corresponding cognitive functions in the brain.

The issues raised above represent only a fraction of the risks worth examining; many others lie beyond the sCope of this piece. More importantly, these four issues are not merely anecdotal observations — each is grounded in an established body of Information Systems (IS) research that predates the widespread adoption of generative AI. Studies of shadow IT, published in *MIS Quarterly Executive* and *Business Information Review*, laid the conceptual groundwork for understanding knowledge leakage through unauthorized technology use well before generative AI emerged. Research on algorithm aversion and algorithm appreciation in *MIS Quarterly* (Turel & Kalhan, 2023) and on cognitive delegation to AI in *Information Systems Research* (Fügener et al., 2022) illuminates the psychological mechanisms behind irrational trust or distrust in AI-generated advice — helping explain why users like those at Deloitte failed to verify AI output. Meanwhile, fake-news research in *MIS Quarterly* (Moravec et al., 2019) and *Information Systems Research* (Moravec et al., 2020) tested platform-level countermeasures to disinformation well before the "Infocalypse" problem was intensified by generative AI. And the cognitive-impact issue traces back to the well-established constructs of cognitive offloading and the "Google effect," studied systematically in *Science* (Sparrow, Liu, & Wegner, 2011) and *Trends in Cognitive Sciences* (Risko & Gilbert, 2016).

What this body of work collectively suggests is that none of these four problems is theoretically new — what generative AI has done is dramatically scale up the speed and magnitude of pre-existing risks. The

research question the IS discipline now needs to answer is therefore not simply *whether* an ethical framework is needed, but *which governance mechanisms actually work* in a generative-AI context. The author hopes the issues discussed here illustrate the urgent need to establish ethical principles, codes of conduct, and governance frameworks for AI use going forward.

Recent work in the *European Journal of Information Systems* has explicitly called on the IS field to engage more systematically with "the dark side of AI" (Mikalef et al., 2022), and the *Journal of Strategic Information Systems* has proposed a research framework for responsible AI governance that remains substantially open for further development (Papagiannidis et al., 2025). In that spirit, this article joins the call for further empirical research on the following specific questions:

1. **Controls against shadow-AI data leakage** — What organizational policies and technical controls allow employees to capture the productivity benefits of generative AI without exposing the organization to confidentiality risk, and how do such policies actually shape real-world usage behavior?
2. **Verification protocols for AI-assisted professional work** — In high-accuracy professions such as consulting, accounting, and law, what verification standards should govern AI-generated or AI-assisted content, and should such standards be set by professional bodies, regulation, or both?
3. **The effectiveness of platform-level anti-fake tools** — Building on System 1/System 2 intervention research (Moravec et al., 2020), how effective are platform-level deepfake- and synthetic-content-detection tools relative to continuously evolving generation techniques?
4. **The long-term cognitive effects of generative-AI reliance on children and adolescents** — Longitudinal research specifically targeting Generation Alpha and Generation Beta is needed, since most current evidence on cognitive offloading is drawn from adult populations and general digital technology rather than generative AI specifically.
5. **Comparative governance models** — platform self-regulation, professional-body oversight, and statutory regulation — which model offers the lowest enforcement cost and highest effectiveness across different industry contexts?

In short, AI has already become part of everyday life, particularly in the context of business, such as business management, marketing, IT infrastructure, and even hiring. It is therefore undeniable that using AI without regard for its boundaries and consequences could cause real harm and broad societal impact. This article aims to underscore that research focused solely on adoption and application is no longer sufficient. The IS discipline — which already possesses the theoretical and methodological foundations for studying human–technology interaction — is uniquely positioned to lead AI governance research in this decade, examining the appropriateness, impact, and sustainability of AI oversight going forward.

Reference:

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2019). Cognitive challenges in human-AI collaboration: Investigating the path towards productive delegation. *Forthcoming, Information Systems Research.* https://dx.doi.org/10.2139/ssrn.3368813

Haag, S., & Eckhardt, A. (2017). Shadow IT. *Business & Information Systems Engineering, 59*(6), 469-473. https://doi.org/10.1007/s12599-017-0497-x

Jeong, D. (2023, March 30). 우려가 현실로…삼성전자, 챗GPT 빗장 풀자마자 '오남용' 속출 [Concerns become reality: Misuse cases erupt as soon as Samsung Electronics lifts ChatGPT ban]. The Economist Korea. https://economist.co.kr/article/view/ecn202303300057?s=31

Mallmann, G. L., Maçada, A. C. G., & Oliveira, M. (2018). The influence of shadow IT usage on knowledge sharing: An exploratory study with IT users. *Business Information Review, 35*(1), 17-28. https://doi.org/10.1177/0266382118760143

McGuirk, R. (2025, October 7). *Deloitte to partially refund Australian government for report with apparent AI-generated errors.* Associated Press. https://apnews.com/article/australia-ai-errors-deloitte-ab54858680ffc4ae6555b31c8fb987f3

McKinsey & Company. (2026). *The state of organizations 2026.* https://www.mckinsey.com/~/media/mckinsey/business%20functions/people%20and%20organizational%20performance/our%20insights/the%20state%20of%20organizations/2026/the-state-of-organizations-2026.pdf

Mikalef, P., Conboy, K., Lundström, J. E., & PopoviČ, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. *European Journal of Information Systems, 31*(3), 257-268. https://doi.org/10.1080/0960085X.2022.2026621

Moravec, P. L., Kim, A., & Dennis, A. R. (2020). Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media. *Information Systems Research, 31*(3), 987-1006. https://doi.org/10.1287/isre.2020.0927

Moravec, P. L., Minas, R. K., & Dennis, A. R. (2019). Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense At All1. MIS quarterly, 43(4), 1343-1360. https://doi.org/10.25300/MISQ/2019/15505

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems, 34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in cognitive sciences, 20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *science, 333*(6043), 776-778. https://doi.org/10.1126/science.1207745

Turel, O., & Kalhan, S. (2023). Prejudiced against the machine? Implicit associations and the transience of algorithm aversion. *Mis Quarterly, 47*(4), 1369-1394. https://doi.org/10.25300/MISQ/2022/17961

Wharton Human-AI Research & GBK Collective. (2025, October 28). *2025 AI adoption report: Gen AI fast-tracks into the enterprise. Knowledge at Wharton*. https://knowledge.wharton.upenn.edu/special-report/2025-ai-adoption-report/

Laddawan Kaewkitipong, Editor-in-Chief