

บทสรุปสำหรับผู้บริหาร

เพื่อให้เกิดประโยชน์สูงสุดจากศักยภาพของปัญญาประดิษฐ์ (AI) และระบบที่ทำงานได้เอง (AS) เราจำเป็นต้องก้าวข้ามการรับรู้และพ้นไปจากการแสวงหาพลังการประมวลผลหรือความสามารถในการแก้ปัญหาที่สูงมากขึ้น

เราจำเป็นต้องทำให้เกิดความมั่นใจว่าเทคโนโลยีเหล่านี้เข้ากันได้กับมนุษย์ ในด้านระบบคุณค่าและหลักการทางจริยธรรม AI/AS ได้ก่อให้เกิดประโยชน์แก่มนุษย์มากกว่าการทำงานได้ผลตามหน้าที่และแก้ปัญหาด้านเทคนิค การทำเช่นนี้จะก่อให้เกิดความไว้วางใจกันระหว่างมนุษย์กับเทคโนโลยีของเราที่จำเป็นต้องการใช้งาน AI/AS อย่างกว้างขวางในชีวิตประจำวันของเรา

“*Eudaimonia*”, หรือ “ชีวิตที่ได้รับพรจากเทพ” ตามที่อาริสโตเติลได้บรรยายไว้ เป็นแนวการปฏิบัติที่กำหนดความหมายของชีวิตที่ดีของมนุษย์ว่าเป็นคุณธรรมสูงสุดของสังคม คำนี้มักแปลเป็นภาษาอังกฤษว่า “ความรุ่งเรือง” ประโยชน์ของ *eudaimonia* เริ่มจากการพิจารณาอย่างตั้งใจ ว่ามิติทางจริยธรรมจะช่วยเรามองเห็นหนทางในการใช้ชีวิตที่ดีได้อย่างไร

การจัดให้การสร้างสรรค์ AI/AS กับระบบคุณค่าของผู้ใช้และสังคมไปในทิศทางเดียวกัน ทำให้เราสามารถจัดลำดับความสำคัญให้แก่การเพิ่มชีวิตที่ดีของมนุษย์ว่าเป็นตัวชี้วัดของการพัฒนาในยุคอัลทริซึมนี้

This document does not represent a position or the views of IEEE but the informed opinions of Committee members providing insights designed to provide expert directional guidance regarding A/IS. This translation is provided for convenience. The English language version of this document is the original and official version of record. In the event of any conflict between the English and translated version (words, terms, phrases, concepts, etc.) the original version of this document ([created in English and available here](#)) governs.

บทสรุปสำหรับผู้บริหาร

เราเป็นใคร

[ความริเริ่มระดับโลกด้านมิติทางจริยธรรมของปัญญาประดิษฐ์และระบบที่ขับเคลื่อนได้เองเป็นโครงการของสถาบันวิศวกรไฟฟ้าและอิเล็กทรอนิกส์ \(IEEE\)](#) ซึ่งเป็นองค์กรวิชาชีพที่มุ่งผลักดันความก้าวหน้าของเทคโนโลยีเพื่อประโยชน์ของมนุษยชาติ มีสมาชิกมากกว่า 400,000 คนในกว่า 160 ประเทศทั่วโลก

ความริเริ่มนี้ได้ให้โอกาสในการนำแนวคิดต่างๆ ในชุมชนของผู้ที่ทำงานด้านปัญญาประดิษฐ์และระบบที่ขับเคลื่อนได้เองเพื่อ [แสวงหาความเห็นร่วมกัน](#) เกี่ยวกับประเด็นที่สำคัญๆ ในเวลานี้

IEEE จัดทำเอกสารเกี่ยวกับการออกแบบที่ถูกต้องหลักจริยธรรม (EAD) การออกแบบที่ถูกต้องหลักจริยธรรม (EAD) จัดทำขึ้นภายใต้หลักการ [Creative Commons Attribution-Non-Commercial 3.0 United States License](#)

ภายใต้การอนุญาตนี้ องค์กรหรือบุคคลสามารถนำเอาส่วนใดส่วนหนึ่งของงานนี้ไปใช้เมื่อใดก็ได้ เป็นที่คาดหวังว่าเนื้อหาของ EAD จะถูกคัดเลือกเพื่อเข้าสู่กระบวนการอย่างเป็นทางการของ IEEE ซึ่งรวมไปถึงการพัฒนามาตรฐานต่างๆ

ความริเริ่มระดับโลกของ IEEE และ EAD จะช่วยเป็นส่วนหนึ่งของความพยายามที่กว้างกว่าที่จัดทำโดย IEEE เพื่อเสริมสร้างการสนทนาในระดับที่กว้างขวางและครอบคลุมเกี่ยวกับจริยศาสตร์ของเทคโนโลยี ซึ่งเรียกกันว่าโปรแกรม [IEEE TechEthics™](#)

บทสรุปสำหรับผู้บริหาร

พันธกิจของควมริเริ่มระดับโลก IEEE

เพื่อให้แน่ใจว่านักเทคโนโลยีทุกคนจะได้รับการศึกษา ฝึกฝนและมีความสามารถเพียงพอในการจัดอันดับความสำคัญของประเด็นทางจริยธรรมในการออกแบบและพัฒนาระบบที่ฉลาดและขับเคลื่อนได้เอง

คำว่า “นักเทคโนโลยี” ในที่นี้หมายถึงใครก็ตามที่ทำงานเกี่ยวข้องกับการวิจัย ออกแบบ ผลิตหรือส่งข้อความเกี่ยวกับ AI/AS รวมไปถึงถึงมหาวิทยาลัย องค์กรและหน่วยงานธุรกิจที่ทำให้เกิดเทคโนโลยีนี้ให้แก่สังคม

เอกสารฉบับนี้เป็นการทำงานร่วมกันของ [นักคิดชั้นนำของโลกจำนวนกว่าหนึ่งร้อยคน](#) ซึ่งมาจากสาขาปัญญาประดิษฐ์ กฎหมาย จริยศาสตร์ ปรัชญาและนโยบายจากสถาบันการศึกษา สถาบันวิจัยวิทยาศาสตร์ ภาครัฐและเอกชน เป้าหมายของเราได้แก่การที่การออกแบบตามหลักจริยธรรมจะให้แก่คิดและข้อเสนอแนะจากผู้เชี่ยวชาญเหล่านี้ ซึ่งเป็นแหล่งอ้างอิงให้แก่ นักเทคโนโลยีด้าน AI/AS ในอนาคตอันใกล้นี้ เพื่อให้บรรลุมิติประสงค์ข้อนี้ ฉบับล่าสุดของ “การออกแบบตามหลักจริยธรรม” จะระบุประเด็น และข้อเสนอแนะต่างๆในสาขาปัญญาประดิษฐ์และระบบที่ขับเคลื่อนได้เอง

เป้าหมายที่สองของควมริเริ่ม IEEE ได้แก่การให้ข้อเสนอแนะเชิงนโยบายเกี่ยวกับมาตรฐานของ IEEE ที่ตั้งอยู่บนการออกแบบอย่างมีจริยธรรม [IEEE P7000](#)
[กระบวนการแบบจำลองสำหรับการพิจารณาประเด็นทางจริยธรรมระหว่างการออกแบบระบบ](#) กระบวนการนี้เป็นโครงการด้านมาตรฐานของ IEEE ที่ได้รับการรับรองและกำลังอยู่ระหว่างการพัฒนา ซึ่งได้แรงบันดาลใจจากควมริเริ่ม โครงการต่อเนื่องอีกสองโครงการ ได้แก่ IEEE p7001TM (โครงการความโปร่งใสของระบบที่ขับเคลื่อนได้เอง) และ IEEE P7002TM (กระบวนการความเป็นส่วนตัวของข้อมูล) ได้รับการอนุมัติแล้ว ซึ่งแสดงให้เห็นถึงการที่ควมริเริ่มมีบทบาทในการกำหนดประเด็นต่างๆในจริยธรรมของ AI/AS

บทสรุปสำหรับผู้บริหาร

โครงสร้างและเนื้อหาของเอกสาร

การออกแบบตามหลักจริยธรรมแบ่งออกเป็นแปดหัวข้อใหญ่ แต่ละหัวข้อกล่าวถึงประเด็นเฉพาะที่เกี่ยวข้องกับ AI/AS ซึ่งได้พิจารณากันอย่างยาวนานในหมู่นักวิชาการและผู้เชี่ยวชาญในความริเริ่ม ประเด็นต่างๆ และข้อเสนอทางนโยบายเหล่านี้จะเรียงลำดับอยู่ในหัวข้อเกี่ยวกับคณะกรรมการ ต่อไปนี้เป็นข้อสรุปของคณะกรรมการต่างๆ รวมถึงประเด็นทางจริยธรรมที่พูดถึงในแต่ละหัวข้อ

1 | หลักการทั่วไป

คณะกรรมการว่าด้วยหลักการทั่วไปพิจารณาเกี่ยวกับมิติทางจริยธรรมในระดับสูงที่ใช้กับประเด็นทางจริยธรรมของ AI/AS ทั้งหมด ประเด็นเหล่านี้ได้แก่

1. การปฏิบัติตามหลักการสูงสุดของสิทธิมนุษยชน
2. การจัดลำดับความสำคัญให้แก่ประโยชน์สูงสุดแก่มวลมนุษยชนและสิ่งแวดล้อมตามธรรมชาติ
3. การเคลื่อนย้ายความเสี่ยงและผลกระทบเชิงลบใน AI/AS เมื่อระบบเหล่านี้ขยายตัวเพิ่มมากขึ้นในระบบสังคมและเทคนิค

ความประสงค์ของคณะกรรมการชุดนี้ได้แก่การที่หลักการประเด็นทางจริยธรรมรวมถึงข้อเสนอแนะต่างๆ ที่พิจารณา จะเป็นพื้นฐานและโครงร่างในการพิจารณาในรายละเอียดต่อไปเกี่ยวกับการกำกับดูแลมิติทางจริยธรรมของการออกแบบ AI/AS

ประเด็นต่างๆ

- เราจะแน่ใจได้อย่างไรว่า AI/AS จะไม่มาละเมิดสิทธิมนุษยชน? (วางหลักการของสิทธิมนุษยชน)
- เราจะทำให้สังคมมั่นใจได้อย่างไรว่าระบบ AI/AS จะมีผู้รับผิดชอบหากเกิดอะไรขึ้น (วางหลักการว่าด้วยการรับผิดชอบ)

- เราจะทำให้สังคมมั่นใจได้อย่างไรว่าระบบ AI/AS มีความโปร่งใส (วางหลักการความโปร่งใส)
- เราจะขยายขอบเขตของประโยชน์ของเทคโนโลยี AI/AS และทำให้ความเสี่ยงของเทคโนโลยีนี้เหลือน้อย

2 | ใส่คุณค่าเข้าไปในระบบปัญญาขับเคลื่อนด้วยตนเอง

การที่จะพัฒนาระบบต่างๆ เกี่ยวกับปัญญาที่ขับเคลื่อนได้ด้วยตนเอง (AIS) ที่เกิดประโยชน์แก่สังคมขึ้นมาได้ ชุมชนนักเทคนิคต้องเข้าใจระบบคุณค่าและใส่ระบบนั้นเข้าไปในระบบปัญญาประดิษฐ์ดังกล่าวนี้ *คณะกรรมการว่าด้วยการใส่คุณค่าเข้าไปในระบบยังมีวัตถุประสงค์ในการใส่ระบบคุณค่าเข้าไปใน AIS โดยมีขั้นตอนสามขั้นเพื่อช่วยงานของนักออกแบบดังนี้*

1. ระบุถึงปทัสฐานและคุณค่าของชุมชนต่างๆ ที่ได้รับผลกระทบจาก AIS

บทสรุปสำหรับผู้บริหาร

2. ทำตามปทัสฐานและคุณค่าของชุมชนเหล่านั้นใน AIS และ
3. ประเมินว่าปทัสฐานกับคุณค่าเหล่านี้ใช้ได้กับมนุษย์และ AIS ทั้งคู่หรือไม่ ภายในชุมชนใดชุมชนหนึ่ง

ประเด็น

- คุณค่าที่จะใส่ไปใน AIS ไม่ใช่คุณค่าสากล แต่เป็นคุณค่าที่เข้ากับชุมชนผู้ใช้เฉพาะแต่ละชุมชน และงานแต่ละงาน
- การมีจริยธรรมมากขึ้นใน AIS มักจะอยู่ภายใต้ปทัสฐานกับคุณค่าจำนวนมาก ซึ่งอาจขัดแย้งกัน
- มีความลำเอียงทางข้อมูลและวิธีคิดคำนวณที่อาจทำให้สมาชิกบางกลุ่มเสียเปรียบ
- เมื่อระบุชุดของปทัสฐาน (เกี่ยวกับบทบาทของ AIS ในชุมชนใดชุมชนหนึ่ง) ได้แล้ว ก็ไม่ชัดเจนว่าปทัสฐานเหล่านี้จะสร้างเข้าไปในโครงสร้างสถาปัตยกรรมการคำนวณได้อย่างไร
- ปทัสฐานที่ใช้อยู่ใน AIS จะต้องเข้ากันได้กับปทัสฐานของชุมชนที่เกี่ยวข้อง.
- หาทางสร้างความไว้วางใจกันระหว่างมนุษย์กับ AIS
- การประเมินความเข้ากันได้ของระบบคุณค่าของ AIS

3 | ระเบียบวิธีในการชี้แนะการวิจัยและออกแบบอย่างมีจริยธรรม

องค์กร AI/AS ในปัจจุบันควรจะทำให้มั่นใจว่าความเป็นอยู่ที่ดีของมนุษย์ การได้รับอำนาจและเสรีภาพจะอยู่ที่แกนกลางของการพัฒนาระบบ AI/AS เพื่อให้สามารถสร้างเครื่องจักรกลที่มีความสามารถทางจริยธรรมเช่นนี้ คณะกรรมการเกี่ยวกับระเบียบวิธีในการชี้แนะการวิจัยและออกแบบอย่างมีจริยธรรม ได้กำหนดประเด็นและข้อเสนอแนะต่างๆที่จะทำให้มั่นใจได้ว่าคุณค่าของมนุษย์ เช่นสิทธิมนุษยชนตามที่ระบุไว้ในปฏิญญาสากลว่าด้วยสิทธิมนุษยชน ได้รับการบรรจุไว้ในระเบียบวิธีของการออกแบบระบบ ระเบียบวิธีที่จัดตรงกับคุณค่าควรจะเป็นจุดสนใจหลักขององค์กรด้าน AI/AS และเป็นระเบียบวิธีที่มุ่งสร้างความก้าวหน้ารุ่งเรืองของมนุษย์ที่ตั้งอยู่บนพื้นฐานอันได้แก่คู่มือทางจริยธรรมเครื่องจักรควรจะได้รับใช้มนุษย์ ไม่ใช่มนุษย์รับใช้เครื่องจักร แนวคิดที่ชอบด้วยจริยธรรมเช่นนี้จะทำให้เกิดความสมดุลระหว่างการเสริมสร้างปัจจัยทางเศรษฐกิจและปัจจัยทางสังคม ซึ่งต่างก็เอื้อต่อการใช้ประโยชน์จาก AI อันจะเป็นผลดีแก่ทั้งธุรกิจและสังคม

ประเด็น

- จริยศาสตร์ไม่ใช่ส่วนหนึ่งของการศึกษาตามหลักสูตร.
- เราจำเป็นต้องมีแบบจำลองของการศึกษาศาสตร์สาขาและสหวัฒนธรรมเพื่อให้เข้าใจประเด็นต่างๆใน AI/AS อย่างครบถ้วน.
- ความจำเป็นที่จะต้องแยกคุณค่าที่แตกต่างกันทางวัฒนธรรมที่ฝังตัวอยู่ในการออกแบบ AI ออกให้เห็นชัด
- การขาดวัฒนธรรมทางจริยธรรมและระบบคุณค่าในการปฏิบัติในอุตสาหกรรม
- การขาดการเป็นผู้นำที่รับรู้ถึงคุณค่า
- การขาดการให้อำนาจที่จะยกประเด็นห่วงใยทางจริยธรรม

บทสรุปสำหรับผู้บริหาร

- การขาดการเป็นเจ้าของและความรับผิดชอบในชุมชนเทคโนโลยี
- ความจำเป็นที่จะต้องรวมเอาผู้มีส่วนได้ส่วนเสียในบริษัทที่ดีที่สุดของ AI/AS
- เอกสารที่เขียนไม่ดีทำให้การออกแบบจริยธรรมเป็นไปได้ยาก
- การติดตามตรวจสอบอัลกอริทึมไม่สม่ำเสมอคงเส้นคงวา
- การขาดองค์หรือสระที่ทำหน้าที่ตรวจสอบ
- การใช้องค์ประกอบจาก “กล่องดำ”

4 | ความปลอดภัยและประโยชน์จาก “ปัญญาประดิษฐ์ทั่วไป” และ “ปัญญาประดิษฐ์อันเป็นเลิศ”

ระบบปัญญาประดิษฐ์ที่มีความสามารถสูงมาก ๆ ในอนาคต (บางครั้งเรียกว่า “ปัญญาประดิษฐ์ทั่วไป” หรือ AGI) อาจจะมีผลกระทบที่ก่อให้เกิดการเปลี่ยนแปลงในโลกในระดับเดียวกับการปฏิวัติการเกษตรหรืออุตสาหกรรม ซึ่งอาจจะนำมาซึ่งความร่ำรวยในโลกอย่างไม่เคยมีมาก่อน คณะกรรมการเกี่ยวกับปัญญาประดิษฐ์ทั่วไปและปัญญาประดิษฐ์อันเป็นเลิศ ได้ให้ข้อเสนอแนะและประเด็นต่างๆ เพื่อให้เกิดความมั่นใจว่าการเปลี่ยนแปลงดังกล่าวนี้จะเป็นไปในทางที่ดี ซึ่งจะมาจากความตั้งใจทำงานจริงของชุมชน AI ที่จะทำให้เป็นไปในทิศทางนั้น

ประเด็น

- เมื่อระบบ AI มีความสามารถมากขึ้น โดยวัดจากความสามารถในการใช้ประโยชน์สูงสุดจากภาระหน้าที่ที่ซับซ้อนมากขึ้นและมีความเป็นตัวของตัวเองมากขึ้นในการทำงานอย่างกว้างขวาง พฤติกรรมที่ไม่ได้คิดไว้ก่อน หรือไม่ได้ตั้งใจจะเป็นอันตรายเพิ่มมากขึ้นมาก
- การติดตั้งระบบความปลอดภัยย้อนหลังให้แก่ระบบ AI ที่มีความสามารถทั่วไปอาจทำได้ยาก
- นักวิจัยและนักพัฒนาจะเผชิญหน้ากับประเด็นด้านความปลอดภัยทางจริยธรรมและทางเทคนิคที่ซับซ้อนมากขึ้นในการพัฒนาและใช้งานระบบ AI ที่เป็นอิสระและมีความสามารถมากขึ้นเรื่อยๆ
- ระบบ AI ในอนาคตอาจจะมีประสิทธิภาพที่จะสร้างผลกระทบให้แก่โลกในระดับเดียวกันกับการปฏิวัติการเกษตรหรืออุตสาหกรรม

5 | ข้อมูลส่วนบุคคลและการควบคุมการเข้าถึงบุคคล

ปัญหาทางสองแฉงที่สำคัญเกี่ยวกับข้อมูลส่วนบุคคลก็คือ *ความไม่สมมาตรของข้อมูล* คณะกรรมการเกี่ยวกับข้อมูลส่วนบุคคลและการเข้าถึงบุคคลได้พิจารณาปัญหานี้และได้ระบุประเด็นทางจริยธรรมรวมไปถึงข้อเสนอข้อเสนอแนะซึ่งแสดงถึงความจำเป็นที่ผู้คนที่ต้องการให้ *คำนิยาม เข้าถึงและจัดการข้อมูลส่วนบุคคล* ของตน ในฐานะที่ข้อมูลเหล่านี้เป็นตัวควบคุมดูแลเอกลักษณ์เฉพาะตนของตนเอง คณะกรรมการรับรู้ว่ามีทั้งทางแก้ที่สมบูรณ์แบบและเครื่องมือดิจิทัลใดๆ ก็อาจจะถูกเจาะระบบได้ อย่างไรก็ตาม คณะกรรมการก็ได้ให้ข้อเสนอแนะเกี่ยวกับการเปิดสิ่งแวดล้อม

บทสรุปสำหรับผู้บริหาร

ทางข้อมูลซึ่งผู้คนสามารถควบคุมความเป็นตัวตนของตนเอง และ
ได้ให้ตัวอย่างของเครื่องมือและแนวทางการปฏิบัติที่สามารถ
ลดล้างความไม่สมมาตรของข้อมูลในอนาคตได้

ประเด็น

- คนๆหนึ่งจะสามารถกำหนดค่านิยมและจัดการข้อมูลส่วนบุคคลของตนเองในยุคอัลกอริทึมได้อย่างไร
- อะไรคือค่านิยมและขอบเขตของข้อมูลข่าวสารที่ทำให้ระบุตัวตนได้
- อะไรคือค่านิยมของการควบคุมเกี่ยวกับข้อมูลส่วนบุคคล
- เราจะให้ค่านิยมใหม่เกี่ยวกับการเข้าถึงข้อมูลเพื่อให้เกียรติแก่ปัจเจกบุคคลได้อย่างไร
- เราจะให้ค่านิยมใหม่เกี่ยวกับคำยินยอมเกี่ยวกับข้อมูลส่วนบุคคลเพื่อให้เกียรติแก่ปัจเจกบุคคลได้อย่างไร
- ข้อมูลที่ดูจะไม่มีคุณลักษณะสำคัญพอที่จะแชร์ กลับสามารถนำมาสร้างการอนุมานที่คนๆหนึ่งไม่อยากจะแชร์
- ผู้จัดการข้อมูลจะทำให้เกิดความมั่นใจว่าผลกระทบ (ทางบวกหรือทางลบ) ของการเข้าถึงและเก็บรวบรวมข้อมูลเป็นกระบวนการที่เปิดเผยชัดเจนแก่ปัจเจกบุคคลคนหนึ่ง เพื่อให้เขาสามารถให้การยินยอมที่มีข้อมูลเพียงพอได้
- บุคคลคนหนึ่งสามารถมีผู้ปกครองที่เป็น AI ส่วนบุคคลหรืออัลกอริทึมได้หรือไม่

6 | การตีกรอบใหม่เกี่ยวกับระบบอาวุธที่ทำงานได้อย่างอิสระ

ระบบที่ทำงานได้อย่างอิสระที่ออกแบบมาเพื่อการก่อให้เกิดความเสียหายแก่ร่างกายมีประเด็นเพิ่มเติมทางจริยธรรมมากขึ้น เมื่อเปรียบเทียบกับอาวุธแบบดั้งเดิม และระบบการทำงานได้อย่างอิสระที่ไม่ได้ออกแบบมาสำหรับการทำร้ายร่างกาย จริยธรรมวิชาชีพเกี่ยวกับเรื่องนี้ควรมีมาตรฐานที่สูงขึ้นเพื่อให้ครอบคลุมประเด็นข้อห่วงใยต่างๆที่กว้างขวางมากขึ้น ในภาพกว้างๆการ “ตีกรอบใหม่” เกี่ยวกับระบบอาวุธที่ทำงานได้อย่างเป็นอิสระ ให้ข้อเสนอแนะว่าองค์กรทางเทคนิคควรจะยอมรับการควบคุมระบบอาวุธเหล่านี้โดยมนุษย์ เพื่อให้เกิดประโยชน์แก่สังคม ยอมรับว่าการตรวจสอบที่มาที่ไปเพื่อตรวจหาการรับผิดชอบ จะทำให้การควบคุมโดยมนุษย์นี้เป็นไปได้จริง ยอมรับว่าผู้ที่สร้างเทคโนโลยีเหล่านี้ขึ้นมา มีความเข้าใจเกี่ยวกับผลกระทบของงานของตนเอง และจรรยาบรรณวิชาชีพควรจะตั้งคำถามถึงเทคโนโลยีที่ตั้งใจออกแบบมาเพื่อทำร้ายกันโดยเฉพาะ

ประเด็น

- จรรยาบรรณวิชาชีพในปัจจุบันที่ของใหม่มาก ซึ่งทำให้ผลงานของสมาชิก รวมไปถึงสิ่งประดิษฐ์และผู้กระทำที่สมาชิก “ได้สร้างขึ้น” ไม่อยู่ภายใต้กรอบจรรยาบรรณเดียวกันกับกรอบที่มายังคับสมาชิก จึงควรมีมาตรการทำให้ผลผลิตและสิ่งประดิษฐ์เหล่านี้อยู่ภายใต้กรอบจรรยาบรรณเดียวกันเท่าที่จะสามารถทำได้
- ความสับสนเกี่ยวกับค่านิยมว่าด้วยมนโศกณศาสำคัญๆในปัญญาประดิษฐ์ ระบบที่ทำงานได้อย่างอิสระ และระบบอาวุธที่ทำงานได้อย่างอิสระ (AWS) ทำให้การอภิปรายอย่างเป็นระบบเกี่ยวกับเรื่องนี้เป็นไปได้ยากลำบาก
- AWS สามารถนำไปใช้ในงานที่เป็นทางลับและหาผู้รับผิดชอบไม่ได้
- มีหนทางหลายทางที่การรับผิดชอบเกี่ยวกับการกระทำของ AWS สามารถทำได้ไปจากแนวทางที่ควรจะเป็นได้

บทสรุปสำหรับผู้บริหาร

- AWS อาจจะทำนายล่วงหน้าไม่ได้ ขึ้นอยู่กับการออกแบบและการใช้งาน ระบบที่เรียนรู้ได้ด้วยตนเองยิ่งทำให้ปัญหาซับซ้อนมากยิ่งขึ้น
- การทำให้การพัฒนา AWS ถูกกฎหมายเป็นการสร้างตัวอย่างที่เป็นอันตรายมากในระดับภูมิยุทธศาสตร์ของโลกในระยะกลาง
- การกันเอาการกำกับดูแลโดยมนุษย์ออกไปจากสมรรถภูมิการรบ อาจจะทำให้การละเมิดสิทธิมนุษยชนอย่างไม่ตั้งใจและทำให้ความตึงเครียดมีมากขึ้น.
- ความหลากหลายของลูกค้าของระบบ AWS ไม่ว่าจะทางตรงหรือทางอ้อม นำไปสู่ภูมิทัศน์ที่ยังยากซับซ้อนเกี่ยวกับการแพร่ขยายและการใช้อาวุธไปในทางที่ผิด
- ตามปกติ ประเภทของอาวุธระบบ AWS จะส่งเสริมให้เกิดการสั่งสมอาวุธกันมากขึ้น และเพิ่มโอกาสในการเกิดความขัดแย้ง
- ยังไม่มีมาตรฐานแน่นอนในการสร้างระบบการตรวจสอบเพื่อให้เกิดความมั่นใจในระบบ AWS
- การเข้าใจขอบเขตทางจริยธรรมระหว่างอาวุธระบบ AWS และอาวุธกึ่งอัตโนมัติ เป็นเรื่องยุ่งยากซับซ้อน

7 | ประเด็นทางเศรษฐกิจและมนุษยธรรม

เทคโนโลยี ระเบียบวิธีวิจัยและระบบต่างๆที่มุ่งลดระดับของการแทรกแซงโดยมนุษย์ในชีวิตประจำวันของเรา กำลังพัฒนาขึ้นอย่างรวดเร็ว และมีแนวโน้มสูงที่จะเปลี่ยนแปลงชีวิตของคนแต่ละคนในหลากหลายมิติ เป้าหมายของคณะกรรมการว่าด้วยประเด็นทางเศรษฐกิจและวัฒนธรรมคือการบ่งชี้ปัจจัยสำคัญในการก่อรูประบบนิเวศน์ของโลกเกี่ยวกับมนุษย์และเทคโนโลยี และอภิปรายถึงผลกระทบทางเศรษฐกิจและมนุษยธรรม แลเสนอโอกาสสำคัญๆในการแก้ปัญหาที่สามารถนำไปใช้ได้ด้วยการปลดล็อกต่างๆเพื่อลดความตึงเครียด เป้าหมายของคณะกรรมการชุดนี้ คือเสนอข้อเสนอนี้เกี่ยวกับข้อห่วงใยเหล่านี้ ซึ่งเกี่ยวเนื่องกับความสัมพันธ์กันระหว่างมนุษย์ สถาบัน และเทคโนโลยีสารสนเทศที่กำลังมีบทบาทเพิ่มมากขึ้น นอกจากนี้ก็มีเป้าหมายในการอำนวยความสะดวกให้แก่การสนทนาข้ามสาขาวิชา ซึ่งการสนทนาจะมีเนื้อหาเพิ่มมากขึ้นจากการมีการคิดร่วมกันในระดับโลกโดยผู้เชี่ยวชาญจากที่ต่างๆที่มีเป้าหมายร่วมกัน

ประเด็น

- การตีความผิดๆเกี่ยวกับ AI/AS ในสื่อมวลชนทำให้ประชาชนเกิดความสับสน
- การทำงานโดยเครื่องจักรมักมองกันว่ามีอยู่แต่ในบริบทของการซื้อขายสินค้าเท่านั้น
- ความซับซ้อนของการจ้างงานกำลังถูกละเลยจากการเข้ามาของหุ่นยนต์และ AI
- การเปลี่ยนแปลงทางเทคโนโลยีกำลังเกิดขึ้นอย่างรวดเร็ว เกิดกว่าที่วิธีการต่างๆที่มีอยู่ในการฝึกฝนอาชีพจะตามทัน
- นโยบายเกี่ยวกับ AI อาจทำให้การผลิตสิ่งของและวิธีการใหม่ๆช้าลง

บทสรุปสำหรับผู้บริหาร

- AI และเทคโนโลยีที่ทำงานอย่างเป็นอิสระมิให้ใช้ไม่เท่ากัน ในภูมิภาคต่างๆในโลก
- การเข้าถึงและการเข้าใจความสำคัญของข้อมูลส่วนบุคคล ยังคงขาดอยู่
- จำนวนของผู้แทนที่ทำงานจริงจากประเทศกำลังพัฒนาการริเริ่มของ IEEE เป็นสิ่งจำเป็น
- การเข้ามาของ AI และระบบที่ทำงานอย่างเป็นอิสระ จะทำให้โครงสร้างทางอำนาจทางเศรษฐกิจและการเมืองระหว่างประเทศพัฒนาแล้วกับประเทศกำลังพัฒนายิ่งเป็นปัญหามากขึ้น

8 | กฎหมาย

การพัฒนา AI/AS ในระยะเริ่มแรกทำให้เกิดปัญหาทางจริยธรรมที่ซับซ้อนหลายประการ ปัญหาเหล่านี้เกือบทั้งหมดจะกลายเป็นข้อท้าทายทางกฎหมาย หรือไม่ก็ก่อให้เกิดปัญหาข้อกฎหมายที่ยู่ยากตามมา คณะกรรมการด้านกฎหมายมีความรู้สึกว่ นักกฎหมายต้องมึงานทำอีกมากในสาขานี้ ซึ่งยังมีนักกฎหมายมาสนใจทำน้อยอยู่มาก แม้ว่าจะมีความจำเป็นอย่างสูง นักกฎหมายจะต้องเป็นส่วนหนึ่งของการอภิปรายว่าด้วยกฎเกณฑ์ การกำกับดูแล กระบวนการออกกฎระเบียบทั้งในระดับชาติและนานาชาติ เพื่อให้ประโยชน์ปริมาณสูงสุดเกิดแก่มวลมนุษย์และโลกของเรา อันเป็นผลจาก AI/AS ที่ได้รับการกำกับดูแลอย่างดีในอนาคต

ประเด็น

- เราจะปรับปรุงการรับผิดชอบและการตรวจสอบในระบบ AI/AS ได้อย่างไร
- เราจะทำให้มั่นใจได้ว่าปัญญาประดิษฐ์จะมีความโปร่งใและเคารพสิทธิของบุคคลได้อย่างไร? ตัวอย่างเช่น รัฐบาลระดับท้องถิ่น ระดับชาติและนานาชาติต่างๆก็ใช้งาน AI ในทางที่ละเมิดสิทธิของพลเมือง ซึ่งควรจะสามารถไว้วางใจรัฐบาลของตนรวมทั้งระบบ AI ว่าจะปกป้องสิทธิของตน
- ระบบ AI จะออกแบบเพื่อให้หลักประกันเรื่องการรับผิดชอบตามกฎหมายหากเกิดอันตรายที่ระบบเป็นฝ่ายก่อให้เกิดขึ้นได้อย่างไร
- ระบบที่ทำงานอย่างเป็นอิสระและมีปัญญา จะออกแบบให้ทำงานในทางที่เคารพบูรณภาพของข้อมูลส่วนบุคคลได้อย่างไร

งานของคณะกรรมการชุดใหม่ของเราจะมีรายละเอียดปรากฏอยู่ในตอนท้ายของ “การออกแบบอย่างมีจริยธรรม”

บทสรุปสำหรับผู้บริหาร

How the Document was Prepared

This document was prepared using an open, collaborative and consensus building approach, following the processes of the Industry Connections program, a program of the IEEE Standards Association. Industry Connections facilitates collaboration among organizations and individuals as they hone and refine their thinking on emerging technology issues, helping to incubate potential new standards activities and standards related products and services.

How to Cite Ethically Aligned Design

Please cite Version 1 of *Ethically Aligned Design* in the following manner:

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. *Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems*, Version 1. IEEE, 2016.
http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

บทสรุปสำหรับผู้บริหาร

Our Appreciation

We wish to express our appreciation for the organizations who have recently contributed research and insights helping to increase awareness around ethical issues and AI/AS, including (but not limited to): [AI Now](#) (White House/New York University); [One Hundred Year Study on Artificial Intelligence](#) (Stanford University); [Preparing for The Future of Artificial Intelligence](#) (U.S. White House/NSTC); [The National Artificial Intelligence Research and Development Strategic Plan](#) (U.S. White House/NSTC); [Robotics and Artificial Intelligence](#) (U.K. House of Commons Science and Technology Committee); Robots and Robotic Devices – [Guide to the Ethical Design and Application of Robots and Robotic Systems](#) (British Standards Institute); [Japan’s Basic Rules for AI Research](#); [Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics](#) (European Parliament); [Éthique de la recherche en robotique](#) (CERNA); [Charta der Digitalen Grundrechte der Europäischen Union](#) ([Charter of the Digital Fundamental Rights of the European Union](#)); and, [Research Priorities for Robust and Beneficial Artificial Intelligence](#) (Future of Life Institute).

We also wish to express our appreciation for the following organizations regarding their seminal efforts regarding AI/AS Ethics, including (but not limited to): [The Association for the Advancement of Artificial Intelligence](#) and their formative work on [AI Ethics](#); [European Association for Artificial Intelligence](#); [ACM Special Interest Group on Artificial Intelligence](#); [The IEEE Robot and Automation Society Committee on Robot Ethics](#); [The IEEE Society on Social Implications of Technology](#); [The Leverhulme Centre for the Future of Intelligence](#); [Allen Institute for Artificial Intelligence](#); [OpenAI](#); [Machine Intelligence Research Institute](#); [Centre for The Study of Existential Risk](#); AI-Austin and, [Partnership on AI to Benefit People and Society](#).

We would also like to acknowledge the contribution of Eileen M. Lach, the General Counsel and Chief Compliance Officer of IEEE, who has reviewed this document in its entirety and affirms the importance of the contribution of The IEEE Global Initiative to the fields of AI/AS ethics.

บทสรุปสำหรับผู้บริหาร

Disclaimers

Ethically Aligned Design is not a code of conduct or a professional code of ethics. Engineers and technologists have well-established codes, and we wish to respectfully recognize the formative precedents surrounding issues of ethics and safety and the professional values these Codes represent. These Codes provide the broad framework for the more focused domain of AI/AS addressed in this document, and it is our hope that the inclusive, consensus-building process around its design will contribute unique value to technologists and society as a whole.

This document is also not a position, or policy statement, or formal report. It is intended to be a working reference tool created in an inclusive process by those in the AI/AS Community prioritizing ethical considerations in their work.

A Note on Affiliations Regarding

Members of The Initiative

The language and views expressed in *Ethically Aligned Design* reflect the individuals who created content for each section of this document. The language and views expressed in this document do not necessarily reflect the Universities or Organizations to which these individuals belong, and should in no way be considered any form of endorsement, implied or otherwise, from these institutions.

This is a first version of *Ethically Aligned Design*. Where [individuals are listed in a Committee](#) it indicates only that they are Members of that Committee. Committee Members may not have achieved final consensus on content in this document because of its versioning format and the consensus-building process of The IEEE Global Initiative for Ethical Consideration in Artificial Intelligence and Autonomous Systems. Content listed by Members in

this or future versions is not an endorsement, implied or otherwise, until formally stated as such.

A Note Regarding Candidate Recommendations in this Document

Ethically Aligned Design is being created via multiple versions that are being iterated over the course of two to three years. The IEEE Global Initiative is following a specific consensus-building process where members contributing content are proposing candidate recommendations so as not to imply these are final recommendations at this time.

Our Membership

Although The IEEE Global Initiative currently has more than one hundred experts from all but one continent involved in our work, most of us come from North America and Europe. We are aware we need to expand our cultural horizons and have more people involved from around the world as we continue to grow our document and our efforts. We are eager for these new voices and perspectives to join our work.

Trademarks and Disclaimers

IEEE believes in good faith that the information in this publication is accurate as of its publication date; such information is subject to change without notice. IEEE is not responsible for any inadvertent errors.

The Institute of Electrical and Electronics Engineers, Incorporated
3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2016 by The Institute of Electrical and Electronics Engineers, Incorporated

บทสรุปสำหรับผู้บริหาร

All rights reserved. Published Month 20xx.

Printed in the United States of America.

IEEE is a registered trademark in the U. S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated.

PDF: ISBN 978-0-7381-xxxx-x STDVxxxxx

Print: ISBN 978-0-7381-xxxx-x STDPDVxxxxx

IEEE prohibits discrimination, harassment, and bullying. For more information, visit <http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html>.

This work is made available under the [Creative Commons Attribution License](#).

To order IEEE Press Publications, call 1-800-678-IEEE.

Find IEEE standards and standards-related product listings at: <http://standards.ieee.org>

Notice and Disclaimer of Liability Concerning the Use of IEEE-SA Industry Connections Documents

This IEEE Standards Association ("IEEE-SA") Industry Connections publication ("Work") is not a consensus standard document. Specifically, this document is NOT AN IEEE STANDARD. Information contained in this Work has been created by, or obtained from, sources believed to be reliable, and reviewed by members of the IEEE-SA Industry Connections activity that produced this Work. IEEE and the IEEE-SA Industry Connections activity members expressly disclaim all warranties (express, implied, and statutory) related to this Work, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; quality, accuracy, effectiveness, currency, or completeness of the Work or content within the Work. In addition, IEEE and the IEEE-SA Industry Connections activity members disclaim any and all conditions relating to: results; and workmanlike effort. This IEEE-SA Industry Connections document is supplied "AS IS" and "WITH ALL FAULTS."

Although the IEEE-SA Industry Connections activity members who have created this Work believe that the information and guidance given in this Work serve as an enhancement to users, all persons must rely upon their own skill and judgment when making use of it. IN NO EVENT SHALL IEEE OR IEEE-SA INDUSTRY CONNECTIONS ACTIVITY MEMBERS BE LIABLE FOR ANY ERRORS OR OMISSIONS OR DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS WORK, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Further, information contained in this Work may be protected by intellectual property rights held by third parties or organizations, and the use of this information may require the user to negotiate with any such rights holders in order to legally acquire the rights to do so, and such rights holders may refuse to grant such rights. Attention is also called to the possibility that implementation of any or all of this Work may require use of subject matter covered by patent rights. By publication of this Work, no position is taken by IEEE with respect to the existence or validity of any patent rights in connection therewith. IEEE is not responsible for identifying patent rights for which a license may be required, or for conducting inquiries into the legal validity or scope of patents claims. Users are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. No commitment to grant licenses under patent rights on a reasonable or non-discriminatory basis has been sought or received from any rights holder. The policies and procedures under which this document was created can be viewed at <http://standards.ieee.org/about/sasb/iccom/>.

This Work is published with the understanding that IEEE and the IEEE-SA Industry Connections activity members are supplying information through this Work, not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought. IEEE is not responsible for the statements and opinions advanced in this Work.