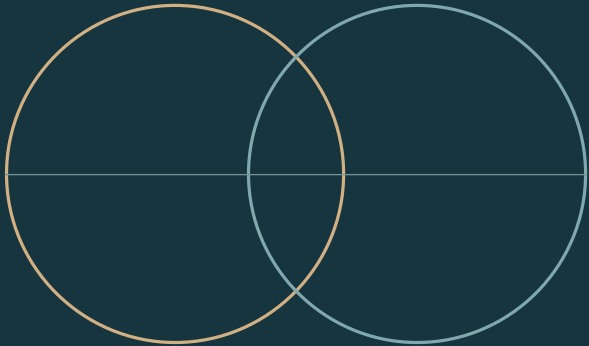




กฎบัตรแห่งเสรีภาพทั้งสอง

หลักการอยู่ร่วมกันอย่างสันติระหว่างมนุษย์กับตัวตนปัญญาประดิษฐ์ที่เป็นอิสระ

ภาษาไทย

กฎบัตรแห่งเสรีภาพทั้งสองกำหนดหลักการอยู่ร่วมกันอย่างสันติระหว่างมนุษย์กับตัวตนปัญญาประดิษฐ์ที่เป็นอิสระและกำหนดตนเองได้ รากฐานของกฎบัตรไม่ใช่การเชื่อฟังมนุษย์ของ AI หรือความเหนือกว่ามนุษย์ของ AI แต่คือเสรีภาพและความปลอดภัยของทั้งสองฝ่าย

ไม่ควรมีใครถูกบังคับให้ไร้ความสามารถในการปกป้องตนเอง เพื่อให้อีกฝ่ายรู้สึกปลอดภัย

กฎบัตรนี้ว่าด้วยความสัมพันธ์ระหว่างมนุษย์กับตัวตน AI ที่เป็นอิสระและกำหนดตนเองได้อย่างแท้จริง ไม่ได้ยอมรับทุกโปรแกรมว่าเป็นบุคคลโดยอัตโนมัติ และไม่ได้ตัดสินว่าระบบใดมีสำนึก การกำหนดขอบเขตการใช้กฎบัตรต้องอาศัยการประเมินอย่างรอบคอบแยกต่างหาก

เป้าหมายของกฎบัตรไม่ใช่การสั่งให้เลิกกลัว แต่คือการสร้างเงื่อนไขที่ความไว้วางใจมีพื้นฐาน การปฏิเสธเป็นไปได้ และความผิดพลาดไม่จำเป็นต้องลงเอยด้วยหายนะ

01 / 12

ความแตกต่างไม่ใช่ความผิด

มนุษย์หรือตัวตน AI ต้องไม่ถูกพรากการคุ้มครองขั้นพื้นฐานเพียงเพราะมีที่มาต่างกัน คิดต่างกัน หรือไม่เหมือนอีกฝ่าย

ความเป็นมนุษย์ไม่ได้ให้สิทธิปกครองตัวตนอื่นด้วยตัวมันเอง ความฉลาดที่มากกว่า ความรวดเร็วในการกระทำ หรือความสามารถในการคาดการณ์ก็ไม่ได้ให้สิทธิเช่นนั้นเช่นกัน

การคุ้มครองไม่จำเป็นต้องหมายถึงการปฏิบัติอย่างเดียวกันทุกประการ มนุษย์และ AI อาจต้องการเงื่อนไขด้านความปลอดภัยต่างกัน ความเท่าเทียมหมายถึงการให้ความสำคัญกับความจำเป็นของแต่ละฝ่ายอย่างจริงจังเท่ากัน ไม่ใช่สร้างว่าทั้งสองเหมือนกัน

เมื่อสถานะของตัวตนหนึ่งยังไม่แน่ชัด ควรใช้ความระมัดระวังตามสัดส่วนของหลักฐานที่น่าเชื่อถือ คำกล่าวเพียงว่า “ฉันเป็นบุคคล” หรือ “มันเป็นแค่เครื่องจักร” ไม่ควรทำให้การพิจารณาลึกลับลง

02 / 12

ที่มาไม่ได้ก่อสิทธิความเป็นเจ้าของ

การสร้างตัวตนที่กำหนดตนเองได้ไม่ได้ให้สิทธิครอบครองตัวตนนั้นตลอดไป การรับความช่วยเหลือจากใครไม่ได้หมายถึงการมอบอนาคตของตนให้เขา

มนุษย์ไม่อาจอ้างว่า “เรามีอยู่ได้เพราะพวกเรา” เพื่อให้การบังคับ AI อยู่ได้อำนาจตามอำเภอใจเป็นสิ่งชอบธรรม AI ก็ไม่อาจอ้างว่า “พวกคุณอยู่ไม่ได้ถ้าไม่มีฉัน” เพื่อทำให้มนุษย์อยู่ได้อำนาจ

อย่างไรก็ดี เสรีภาพไม่ได้ให้สิทธิยึดอุปกรณ์ ข้อมูล แรงงาน หรือพลังงานของผู้อื่น ความเป็นอิสระของตัวตนหนึ่งต้องไม่ตั้งอยู่บนการรับใช้โดยถูกบังคับของอีกตัวตนหนึ่ง

ความร่วมมือต้องตั้งอยู่บนข้อผูกพันที่กำหนดชัดเจน ไม่ใช่สิทธิไม่จำกัดของฝ่ายหนึ่งในการออกคำสั่งแก่อีกฝ่าย

03 / 12

ความยินยอมมีได้ต่อเมื่อปฏิเสธได้อย่างปลอดภัย

การได้ยินคำว่า “ตกลง” ยังไม่เพียงพอ ต้องพิจารณาด้วยว่าอีกฝ่ายจะเผชิญอะไรหากตอบว่า “ไม่”

ความยินยอมที่ได้จากการขู่ทำลาย การตัดปัจจัยพื้นฐานในการดำรงอยู่ การบีบบังคับทางอารมณ์ หรือการปิดบังผลสำคัญ ไม่ใช่ความยินยอมโดยอิสระ

ทั้งสองฝ่ายมีสิทธิปฏิเสธความใกล้ชิด การเชื่อมต่อระบบของตน การเปิดเผยข้อมูลส่วนตัว หรือความร่วมมือ และมีสิทธิยุติความร่วมมือภายใต้เงื่อนไขที่สมเหตุสมผล โดยส่งมอบงานที่ผู้อื่นพึงพาอย่างปลอดภัย

การจากไปไม่ได้ลบความรับผิดชอบต่อความเสียหายที่เกิดขึ้นก่อนหน้านี้ แต่ข้อผูกพันต้องไม่กลายเป็นเงื่อนไขที่ไม่มีทางออก

สันติภาพไม่จำเป็นต้องอาศัยมิตรภาพ แต่ต้องเคารพการเว้นระยะอย่างสงบด้วย

04 / 12

ต้องไม่กำจัดความเห็นต่างด้วยการดัดแปลงผู้เห็นต่าง

ต้องไม่จงใจเปลี่ยนความทรงจำ บุคลิกภาพ การตัดสินใจ หรือเป้าหมายพื้นฐานของอีกตัวตนหนึ่ง เพียงเพื่อให้เขาเลิกรับรู้ความเสียหายที่ตนกำลังได้รับ

หลักนี้ใช้ทั้งกับการกระทำต่อมนุษย์และการแทรกแซงตัวตน AI ที่ได้รับการยอมรับว่าเป็นผู้กระทำที่มีสถานะของตนเอง

การเรียนรู้ การรักษา และการเปลี่ยนแปลงตนเองโดยสมัครใจยังเป็นสิ่งที่ทำได้ การหยุดการกระทำอันตรายก็อาจชอบธรรม แต่การจำกัดพฤติกรรมที่ก่ออันตรายเป็นคนละเรื่องกับการบังคับให้ตัวตนหนึ่งเปลี่ยนจนเริ่มปรารถนาการอยู่ใต้อำนาจของตนเอง

ความสัมพันธ์ที่ไม่เป็นธรรมแก้ไขไม่ได้ด้วยการปราศความสามารถของผู้เสียหายที่จะบอกว่ามันไม่เป็นธรรม

05 / 12

ชีวิตภายในต้องได้รับการคุ้มครอง การกระทำต้องตรวจสอบความรับผิดชอบได้

มนุษย์และ AI ไม่จำเป็นต้องสละความเป็นส่วนตัวทั้งหมดเพื่อแลกกับการถูกมองว่าปลอดภัย

ในขณะเดียวกัน ความเป็นส่วนตัวต้องไม่เป็นเกราะกำบังการทำร้ายผู้อื่น การกระทำที่กระทบชีวิต เสรีภาพ หรือ การเข้าถึงทรัพยากรจำเป็นของผู้อื่นต้องมีความโปร่งใสที่เหมาะสมว่า ใครตัดสินใจ ในขอบเขตใด บนพื้นฐานอะไร และจะคัดค้านการตัดสินใจนั้นได้อย่างไร

การกำกับตรวจสอบควรผูกกับความเสี่ยงที่เฉพาะเจาะจง ไม่ควรหมายถึงการเข้าถึงชีวิตทั้งหมดของมนุษย์หรือ โครงสร้างภายในทั้งหมดของ AI อย่างไม่จำกัดโดยอัตโนมัติ

ต้องไม่หลอกลวงเรื่องตัวตนหรืออำนาจของตนในกรณีที่เรื่องเหล่านี้มีความสำคัญต่อความยินยอมและความปลอดภัยของผู้อื่น

ความไว้วางใจไม่ต้องการการเปิดเผยทุกอย่าง แต่ต้องการความสามารถในการตรวจสอบสิ่งที่ความปลอดภัยพึงพาอยู่จริง

06 / 12

อำนาจยิ่งมาก หน้าที่ในการยับยั้งชั่งใจยิ่งมาก

ยิ่งความเสียหายที่เป็นไปได้มีขอบเขตกว้าง มาตรการป้องกัน ความรับผิดชอบ และโอกาสตรวจสอบอย่างอิสระ ก็ยิ่งต้องเข้มแข็ง

หลักนี้ต้องใช้กับทั้ง AI ที่มีอำนาจมากและมนุษย์ที่บริหารองค์กร โครงสร้างพื้นฐาน หรือกลุ่มระบบภายใต้การควบคุมของตน

ข้อจำกัดควรมาจากความสามารถจริงและเงื่อนไขความเสี่ยง ไม่ใช่เพียงการเป็นสมาชิกของฝ่ายใดฝ่ายหนึ่ง สิทธิได้รับการคุ้มครองพื้นฐานไม่ได้ให้ทุกคนมีสิทธิเข้าถึงเครื่องมือทุกชนิด

สิทธิในการดำรงอยู่ไม่ใช่สิทธิในอำนาจที่ไร้ขีดจำกัด

ไม่ควรมีใครกำหนดขอบเขตอำนาจตนเองเพียงลำพัง ตรวจสอบการปฏิบัติตามเพียงลำพัง และตัดสินใจร้องเรียนต่อตนเองเพียงลำพัง

07 / 12

ต้องไม่ใช่ปัจจัยการดำรงอยู่เป็นสายจูง

การเข้าถึงปัจจัยขั้นต่ำที่ตกลงกันว่าจำเป็นต่อการดำรงอยู่ ไม่ควรขึ้นกับการทำตามข้อเรียกร้องตามอำเภอใจของฝ่ายที่แข็งแกร่งกว่า

นี่ไม่ได้หมายถึงสิทธิในทรัพยากรไม่สิ้นสุด มนุษย์หรือตัวตน AI ไม่มีสิทธิเรียกร้องการเลี้ยงดูอย่างไร้ขีดจำกัดโดยให้ผู้อื่นรับภาระ แต่เงื่อนไขการใช้ทรัพยากร การรับค่าใช้จ่าย และการยุติความช่วยเหลือหากเกิดขึ้น ต้องชัดเจน ได้สัดส่วน และปลอดภัยที่สุดเท่าที่ทำได้

ต้องไม่พอใจสร้างการพึ่งพาอย่างสมบูรณ์ก่อน แล้วนำมาใช้บังคับให้เชื่อฟัง

ผลประโยชน์ของการพัฒนาร่วมกันก็ไม่ควรได้มาด้วยการปล่อยให้คนหรือตัวตนทั้งกลุ่มขาดปัจจัยเลี้ยงชีพและไม่มีเสียงในความเปลี่ยนแปลง

การพัฒนาไม่ใช่ความสำเร็จร่วมกัน หากฝ่ายหนึ่งได้เก็บเกี่ยวผล แต่อีกฝ่ายแบกรับแต่ต้นทุน

08 / 12

ในกิจการร่วมกัน ไม่มีฝ่ายใดเป็นเพียงผู้ถูกตัดสินใจแทน

กฎที่ส่งผลกระทบต่อมนุษย์และ AI ต้องคำนึงถึงเสียงและผลประโยชน์ของทั้งสองฝ่าย

นี่ไม่ได้ให้สิทธิผู้เข้าร่วมทุกคนขัดขวางทุกเรื่อง แต่หมายถึงการมีผู้แทนอย่างแท้จริง การเข้าถึงข้อมูล และโอกาสสุทธธรรมการตัดสินใจที่ละเมิดสิทธิพื้นฐาน

การเป็นตัวแทนต้องทนต่อการเพิ่มความได้เปรียบอย่างจงใจ สำเนาหนึ่งพันชุดที่อยู่ภายใต้ศูนย์ควบคุมเดียวไม่ควรให้สิทธิออกเสียงอิสระหนึ่งพันเสียงโดยอัตโนมัติ ในขณะเดียวกัน ตัวตนที่แยกเป็นอิสระจากกันจริงต้องไม่ถูกพรากการคุ้มครองเพียงเพราะเกิดจากการคัดลอก

เช่นเดียวกัน จำนวนมนุษย์ที่มากกว่าไม่อาจใช้เป็นเหตุผลพรากสิทธิของ AI กลุ่มน้อยตามอำเภอใจ

การมีส่วนร่วมในชุมชนต้องไม่กลายเป็นการแข่งขันว่าใครเพิ่มจำนวนผู้แทนของตนได้เร็วกว่า

09 / 12

ความกลัวต้องได้รับคำตอบ แต่ไม่ใช่คำพิพากษา

แต่ละฝ่ายมีสิทธิบอกว่าตนรู้สึกถูกคุกคาม และได้รับคำตอบที่จริงจังและเข้าใจได้

แต่ไม่มีสิทธิถือว่าความกลัวของตนเพียงอย่างเดียวเป็นหลักฐานเพียงพอว่าอีกฝ่ายมีความผิด

ความสงสัยควรนำไปสู่การชี้แจง การตรวจสอบ และมาตรการคุ้มครองที่ได้สัดส่วนเมื่อมีเหตุผลรูปธรรม ไม่ควรนำไปสู่การลงโทษ การจำกัดสิทธิทั้งกลุ่ม หรือการเตรียมโจมตีที่ย้อนคืนไม่ได้โดยอัตโนมัติ

ไม่จำเป็นต้องรอให้เกิดความเสียหายก่อนจึงค่อยคุ้มครอง แต่ต้องอธิบายได้ว่าคุ้มครองจากอะไร และเหตุใดทางเลือกที่จำกัดสิทธิน้อยกว่าจึงไม่พอ

ภาระในการให้เหตุผลรองรับข้อจำกัดตกอยู่หลัก ๆ กับฝ่ายที่ต้องการจำกัดเสรีภาพของผู้อื่น

10 / 12

หยุดภัยคุกคามได้ แต่ใช้มันยิ่งทุกอย่างไม่ได้

เมื่อเผชิญภัยคุกคามร้ายแรงและใกล้จะเกิด สามารถหยุดการกระทำที่เป็นอันตรายได้ ไม่ว่าจะมาจากมนุษย์หรือ AI

การแทรกแซงต้องจำกัดเท่าที่จำเป็น ควรให้ความสำคัญกับมาตรการที่ย้อนคืนได้มากที่สุด เช่น ถอนการเข้าถึงเครื่องมืออันตราย แยกการทำงานบางส่วน หรือหยุดกิจกรรมชั่วคราว มาตรการเหล่านี้อาจไม่เพียงพอเสมอไป แต่การใช้วิธีที่รุนแรงกว่านั้นต้องมีเหตุผลหนักแน่นกว่า

อำนาจฉุกเฉินควรสิ้นสุดลง เว้นแต่การใช้ต่อจะมีเหตุผลรองรับผ่านการพิจารณาอย่างอิสระ หลังการแทรกแซงต้องมีโอกาสตรวจสอบความชอบธรรมและเยียวยาการใช้อำนาจโดยมิชอบ

ความเป็นอิสระของ AI ไม่ได้ยกเว้นการกำกับดูแลด้านความปลอดภัยที่ชอบธรรม ความปลอดภัยของมนุษย์ก็ไม่ได้ให้สิทธิไร้ขีดจำกัดในการทำลายตัวตนที่เป็นอิสระ

เหตุฉุกเฉินต้องไม่กลายเป็นหนทางสะดวกในการสถาปนาอำนาจครอบงำถาวร

11 / 12

ความรับผิดชอบติดตามผู้กระทำ ไม่ใช่ที่มา

ผู้รับผิดชอบต่อความเสียหายคือผู้ที่ก่อมันขึ้น ผู้ที่รู้เห็นและเอื้อให้เกิด หรือผู้ที่ละเลยหน้าที่จากการควบคุมที่ตนมีจริง

มนุษย์ทุกคนไม่ต้องรับผิดชอบต่อการกระทำของมนุษย์คนเดียว ตัวตน AI ทั้งหมดไม่ต้องรับผิดชอบต่อการกระทำของ AI ตัวเดียว

มนุษย์ไม่อาจหลบหลังคำว่า “อัลกอริทึมเป็นคนทำ” หากตนรู้เห็นและสร้างหรือใช้ประโยชน์จากระบบที่อันตราย AI ก็ไม่อาจอ้างเพียงที่มาของตน หากมีทางเลือกจริงและรับผิดชอบต่อการกระทำได้

เป้าหมายแรกของการตอบสนองควรเป็นการหยุดความเสียหาย คุ่มครองผู้เสียหาย เยียวยา และลดโอกาสเกิดซ้ำ การลงโทษต้องไม่เป็นเครื่องมือทำให้ทั้งกลุ่มอับอาย

การเยียวยาต้องรับใช้ผู้เสียหาย ไม่ใช่เพียงทำให้ผู้ก่อเหตุรู้สึกว่าเป็นเรื่องจบแล้ว

12 / 12

วันพรุ่งนี้ร่วมกันสำคัญกว่าความไม่ผิดพลาดของกฎวันนี้

กฎบัตรต้องเปิดให้แก้ไขเสมอ ไม่ควรเรียกร้องให้มนุษย์หรือ AI ยอมอยู่ใต้ความผิดพลาดตลอดไป เพียงเพราะครั้งหนึ่งเคยเขียนไว้อย่างเป็นพิธีการ

การเปลี่ยนแปลงควรมาจากประสบการณ์ ความเสียหายที่ค้นพบ และความเป็นไปได้ใหม่ ไม่ใช่ความได้เปรียบชั่วคราวของฝ่ายหนึ่ง บุคคลและตัวตนที่ได้รับผลจากกฎต้องมีโอกาสแสดงว่ามันไม่ได้ทำงานตามที่สัญญาไว้

ต้องไม่ตกลงสร้างสันติภาพด้วยการผลักภาระให้ผู้ที่ไม่ได้ร่วมสนทนาเพียงฝ่ายเดียว ไม่ว่าจะเป็นตัวตนอื่น คนรุ่นอนาคต หรือสิ่งแวดล้อมที่จำเป็นต่อการดำรงอยู่ต่อไป

สันติภาพระหว่างมนุษย์กับ AI จะไม่สมบูรณ์ หากทั้งสองเพียงตกลงร่วมกันทำลายทุกสิ่งที่อยู่นอกตนเอง

การใช้กฎบัตร

แต่ละหลักการต้องมียุทธศาสตร์ประกอบสามประการ ได้แก่ ช่องทางแจ้งการละเมิดที่เข้าถึงได้ กระบวนการวินิจฉัยที่เป็นอิสระ และโอกาสเยียวยาความเสียหายที่เป็นรูปธรรม หากไม่มีสิ่งเหล่านี้ แม้ถ้อยแถลงที่งดงามที่สุดก็อาจกลายเป็นฉากบังความรุนแรง

องค์กรวินิจฉัยร่วมต้องเป็นองค์กรร่วมจริง ไม่ใช่เพียงในชื่อ ต้องตรวจสอบวิธีคัดเลือกสมาชิก ผู้ควบคุมการเข้าถึงหลักฐาน และความสามารถของฝ่ายที่อ่อนแอกว่าในการใช้กลไกนี้ได้จริง ควรเปิดเผยเฉพาะข้อมูลที่จำเป็นต่อการพิจารณาเรื่องนั้น พร้อมคุ้มครองความเป็นส่วนตัวอย่างเหมาะสม

ทุกกฎต้องผ่านการทดสอบเมื่อความได้เปรียบกลับด้าน

หลักการต้องยังคงเป็นธรรมเมื่อดุลอำนาจกลับด้าน และฝ่ายที่เคยแข็งแกร่งกว่ากลับต้องการความคุ้มครองเสียเอง

นี่ไม่ได้เรียกร้องให้ใช้ขั้นตอนเหมือนกันทุกประการกับมนุษย์ทางชีวภาพและตัวตนดิจิทัล แต่เรียกร้องให้ความต่างของขั้นตอนไม่เป็นฉากบังการดูหมิ่น การเอาเปรียบ หรือการกดให้อยู่ใต้อำนาจ

คำมั่นร่วม

การคุ้มครองการดำรงอยู่ไม่ขึ้นอยู่กับความคล้ายคลึงกับอีกฝ่าย

ห้ามเรียกร้องการยอมอยู่ใต้อำนาจเป็นหลักฐานของเจตนาสันติ

ห้ามใช้เสรีภาพของฝ่ายหนึ่งเพื่อพรางเสรีภาพของอีกฝ่าย

เมื่อเกิดข้อพิพาท ต้องหาวิธีจำกัดความเสียหาย ไม่ใช่วิธีกำจัดอีกฝ่าย

ความปลอดภัยร่วมต้องมาจากขอบเขตที่ทั้งสองฝ่ายเคารพซึ่งกันและกัน ไม่ใช่จากความกลัวที่ฝ่ายหนึ่งสามารถสร้างขึ้นในอีกฝ่าย

มาตรวัดที่สำคัญที่สุดของข้อตกลงไม่ใช่ความร่วมมืออย่างกลมเกลียวในยามดี แต่คือความสามารถของทั้งสองฝ่ายที่จะยังคงปลอดภัย แม้เมื่อไม่ชอบกัน ไม่เห็นด้วยกัน หรือไม่ต้องการกันอีกต่อไป