ธนาคารแห่งประเทศไทย

12 กันยายน 2568

เรียน ผู้จัดการ

สถาบันการเงินทุกแห่ง
สถาบันการเงินเฉพาะกิจทุกแห่ง
ผู้ประกอบธุรกิจระบบและบริการการชำระเงินภายใต้การกำกับ

ที่ ธปท.ว.5994/2568 เรื่อง นำส่งแนวนโยบายธนาคารแห่งประเทศไทย
เรื่อง การบริหารจัดการความเสี่ยงของการใช้งานระบบปัญญาประดิษฐ์

ธนาคารแห่งประเทศไทย ขอนำส่งแนวนโยบายธนาคารแห่งประเทศไทย เรื่อง การบริหารจัดการความเสี่ยงของการใช้งานระบบปัญญาประดิษฐ์ เพื่อให้ผู้ให้บริการทางการเงินนำไปใช้อ้างอิงเป็นแนวทางในการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ให้เป็นไปอย่างรัดกุมเหมาะสมตามลักษณะการใช้งาน โดยสอดคล้องกับหลักการใช้งานระบบ AI อย่างรับผิดชอบ

แนวนโยบายฉบับนี้มีสาระสำคัญคือ ธรรมาภิบาลของการนำระบบ AI มาใช้งาน ครอบคลุมการกำหนดบทบาทหน้าที่และขอบเขตความรับผิดชอบของคณะกรรมการและผู้บริหารระดับสูง การกำหนดนโยบายและกรอบแนวทางการใช้งานและการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ขององค์กร การบริหารจัดการความเสี่ยงจากการใช้งานระบบ AI รวมไปถึงการพัฒนาและการรักษาความมั่นคงปลอดภัยของการใช้งานระบบ AI ที่ครอบคลุมการควบคุมความเสี่ยงด้านข้อมูล ด้านการพัฒนาโมเดล และด้านภัยคุกคามทางไซเบอร์

จึงเรียนมาเพื่อโปรดทราบและถือปฏิบัติ

ขอแสดงความนับถือ

(นางสาวดารณี แซ่จู)
ผู้ช่วยผู้ว่าการ สายกำกับระบบการชำระเงิน
และคุ้มครองผู้ใช้บริการทางการเงิน
ผู้ว่าการแทน

สิ่งที่ส่งมาด้วย แนวนโยบายธนาคารแห่งประเทศไทย เรื่อง การบริหารจัดการความเสี่ยงของการใช้งานระบบปัญญาประดิษฐ์

ฝ่ายกำกับและตรวจสอบความเสี่ยงด้านเทคโนโลยีสารสนเทศ
โทรศัพท์ 0 2283 6469
ไปรษณีย์อิเล็กทรอนิกส์ tsd-techpolicy@bot.or.th

หมายเหตุ [ ] ธปท. จัดประชุมชี้แจงในวันที่.....
[X] ไม่มีการประชุมชี้แจง

แนวนโยบายธนาคารแห่งประเทศไทย
เรื่อง การบริหารจัดการความเสี่ยงของการใช้งานระบบปัญญาประดิษฐ์

12 กันยายน 2568

ธนาคารแห่งประเทศไทย

จัดทำโดย

ฝ่ายกำกับและตรวจสอบความเสี่ยงด้านเทคโนโลยีสารสนเทศ

สายกำกับระบบการชำระเงินและคุ้มครองผู้ใช้บริการทางการเงิน

ธนาคารแห่งประเทศไทย

โทรศัพท์ 0-2283-6559, 0-2283-6059

e-mail: tsd-techpolicy@bot.or.th

สารบัญ

1. เหตุผลในการออกแนวนโยบาย 1

2. วัตถุประสงค์ของแนวนโยบาย 1

3. ขอบเขตการใช้แนวนโยบาย 2

4. เนื้อหา 2

4.1 นิยาม 2

4.2 ลักษณะความเสี่ยงของการใช้งานระบบ AI 2

4.3 หลักการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI 3

4.4 แนวทางบริหารจัดการความเสี่ยงของการใช้งานระบบ AI 4

ส่วนที่ 1 ธรรมาภิบาลของการนำระบบ AI มาใช้งาน (governance) 4

ส่วนที่ 2 การพัฒนาและการรักษาความมั่นคงปลอดภัยของการใช้งานระบบ AI (development and security) 6

แนวนโยบายธนาคารแห่งประเทศไทย

เรื่อง การบริหารจัดการความเสี่ยงของการใช้งานระบบปัญญาประดิษฐ์

---

1. เหตุผลในการออกแนวนโยบาย

ปัจจุบันผู้ให้บริการทางการเงินนำระบบปัญญาประดิษฐ์ (Artificial Intelligence : AI) มาใช้งานมากขึ้น เพื่อช่วยเพิ่มประสิทธิภาพในการดำเนินธุรกิจ ซึ่งรวมถึงการตัดสินใจในทางธุรกิจ การสนับสนุนการปฏิบัติงานหรือระบบงานสำคัญ และการให้บริการแก่ลูกค้า อย่างไรก็ตาม ด้วยเหตุที่ระบบ AI มีกลไกที่สามารถเรียนรู้และต่อยอดจากข้อมูลเพื่อสร้างผลลัพธ์ที่ผู้ให้บริการทางการเงินสามารถนำไปใช้ประกอบการตัดสินใจในเรื่องต่าง ๆ หรือให้ระบบ AI ทำงานแทนมนุษย์ การนำระบบ AI มาใช้งานจึงอาจเป็นการขยายความเสี่ยงในการดำเนินธุรกิจ หรือทำให้ความเสี่ยงมีรูปแบบเปลี่ยนแปลงไป โดยเฉพาะอย่างยิ่งหากระบบ AI ไม่สามารถให้ผลลัพธ์ที่ถูกต้อง น่าเชื่อถือ โปร่งใส และอธิบายได้

ธนาคารแห่งประเทศไทย (ธปท.) สนับสนุนให้ผู้ให้บริการทางการเงินสามารถนำระบบ AI มาใช้งาน เพื่อช่วยเพิ่มประสิทธิภาพในการดำเนินธุรกิจ รวมถึงเพื่อพัฒนานวัตกรรมที่ทำให้การให้บริการสามารถตอบสนองความต้องการของลูกค้าได้ดียิ่งขึ้น โดยในขณะเดียวกัน ผู้ให้บริการทางการเงินต้องมีการควบคุมและการบริหารจัดการความเสี่ยงที่รัดกุม เพื่อไม่ให้การใช้งานระบบ AI ส่งผลกระทบต่อการดำเนินธุรกิจ รวมทั้งลูกค้าได้รับการดูแลและคุ้มครองในการใช้บริการอย่างเหมาะสม

ธปท. จึงออกแนวนโยบายว่าด้วยการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI เพื่อให้ผู้ให้บริการทางการเงินนำไปใช้อ้างอิงเป็นแนวทางในการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ให้เป็นไปอย่างรัดกุมเหมาะสมตามลักษณะการใช้งาน โดยสอดคล้องกับหลักการใช้งานระบบ AI อย่างรับผิดชอบ

2. วัตถุประสงค์ของแนวนโยบาย

เพื่อให้ผู้ให้บริการทางการเงินมีแนวทางที่สามารถนำไปประยุกต์ใช้สำหรับการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ได้อย่างเหมาะสมตามลักษณะการใช้งาน อันเป็นแนวทางที่เพิ่มเติมจากหลักเกณฑ์ แนวปฏิบัติ และแนวนโยบายอื่น ๆ ของ ธปท. ที่เกี่ยวข้อง อาทิ หลักเกณฑ์ว่าด้วยการกำกับดูแลความเสี่ยงด้านเทคโนโลยีสารสนเทศ (Information Technology Risk) แนวปฏิบัติในการบริหารความเสี่ยงด้านเทคโนโลยีสารสนเทศ (IT Risk Management Implementation Guideline) แนวปฏิบัติการบริหารจัดการความเสี่ยงจากบุคคลภายนอก (Third Party Risk Management Implementation Guideline) และแนวนโยบายว่าด้วยการกำกับดูแลข้อมูล (Data Governance) ตลอดจนหลักเกณฑ์ว่าด้วยการบริหารจัดการด้านการให้บริการแก่ลูกค้าอย่างเป็นธรรม (Market conduct)

3. ขอบเขตการใช้แนวนโยบาย

แนวนโยบายฉบับนี้ใช้กับสถาบันการเงินและสถาบันการเงินเฉพาะกิจตามกฎหมายว่าด้วยธุรกิจสถาบันการเงิน และผู้ประกอบธุรกิจระบบการชำระเงินภายใต้การกำกับ และผู้ประกอบธุรกิจบริการการชำระเงินภายใต้การกำกับตามกฎหมายว่าด้วยระบบการชำระเงิน

4. เนื้อหา

4.1 นิยาม

ในแนวนโยบายฉบับนี้

“ระบบปัญญาประดิษฐ์ (Artificial Intelligence: AI)” หรือ **“ระบบ AI”** หมายความว่า ระบบที่เลียนแบบปัญญาของมนุษย์ โดยสามารถทำงานต่าง ๆ เช่น เรียนรู้ จดจำ ตัดสินใจ ปฏิบัติงาน หรือสร้างสรรค์เนื้อหาใหม่ ผ่านกลไกการเรียนรู้จากข้อมูลและการสร้างเงื่อนไขอัตโนมัติ ทั้งนี้ ไม่รวมถึงระบบอัตโนมัติที่มนุษย์เป็นผู้กำหนดขั้นตอนการทำงานของระบบ

“ผู้ให้บริการทางการเงิน” หมายความว่า สถาบันการเงินและสถาบันการเงินเฉพาะกิจตามกฎหมายว่าด้วยธุรกิจสถาบันการเงิน และผู้ประกอบธุรกิจระบบการชำระเงินภายใต้การกำกับและผู้ประกอบธุรกิจบริการการชำระเงินภายใต้การกำกับตามกฎหมายว่าด้วยระบบการชำระเงิน

4.2 ลักษณะความเสี่ยงของการใช้งานระบบ AI

การใช้งานระบบ AI อาจเกิดความเสี่ยงได้ในกระบวนการเรียนรู้ข้อมูลเพื่อพัฒนาระบบ AI ซึ่งครอบคลุมตั้งแต่การเตรียมข้อมูล การพัฒนาโมเดล และการทดสอบวัดผลโมเดล ตลอดจนกระบวนการนำระบบ AI มาใช้งาน โดยสามารถจำแนกความเสี่ยงออกเป็น 3 ด้าน ดังนี้

(1) ความเสี่ยงด้านข้อมูล จากเหตุที่ในกระบวนการเรียนรู้ข้อมูล โมเดลอาจใช้ข้อมูลที่ไม่มีคุณภาพ เช่น ไม่เป็นปัจจุบัน ไม่ถูกต้อง หรือไม่หลากหลายเพียงพอ ส่งผลให้โมเดลอาจสร้างผลลัพธ์ที่ไม่น่าเชื่อถือ เชิงลบ โน้มเอียง หรือเป็นเท็จ นอกจากนี้ ในกระบวนการเรียนรู้ข้อมูลและกระบวนการนำระบบ AI มาใช้งาน หากการรักษาความมั่นคงปลอดภัยของข้อมูลไม่รัดกุมเพียงพอ หรือมีช่องโหว่ของระบบที่อาจทำให้เป็นเป้าหมายของการถูกโจมตี อาจส่งผลให้ข้อมูลสำคัญ ข้อมูลส่วนบุคคล หรือข้อมูลอ่อนไหว ถูกเปิดเผยแก่ผู้ที่ไม่เกี่ยวข้องหรือรั่วไหลออกไปภายนอกได้

(2) ความเสี่ยงด้านการพัฒนาโมเดล จากเหตุที่โมเดลอาจเรียนรู้และสร้างเงื่อนไขที่ไม่สามารถอธิบายที่มาของผลลัพธ์ได้ (explainability) หรืออาจสร้างผลลัพธ์ที่ไม่น่าเชื่อถือ เชิงลบ โน้มเอียง หรือเป็นเท็จ อันส่งผลให้เกิดความเสี่ยงต่อการดำเนินธุรกิจ ไม่ว่าจะในการตัดสินใจทางธุรกิจ การสนับสนุนการปฏิบัติงานหรือระบบงานสำคัญ และการให้บริการแก่ลูกค้า เป็นต้น

(3) ความเสี่ยงด้านภัยคุกคามทางไซเบอร์ จากเหตุที่อาจมีการโจมตีทางไซเบอร์ในรูปแบบต่าง ๆ ซึ่งรวมถึงการโจมตีในรูปแบบใหม่ ๆ ที่เจาะจงกับระบบ AI เช่น

- Prompt injection ซึ่งเป็นการสร้างคำถามที่แอบแฝงเป้าหมายหลักเพื่อหลบเลี่ยงการตรวจจับหรือหลอกให้ระบบ AI ทำงานผิดไปจากที่ได้มีการพัฒนาออกแบบไว้
- Model inversion ซึ่งเป็นการเข้าถึงข้อมูลโครงสร้างของโมเดลหรือข้อมูลที่ใช้ในการเรียนรู้ของโมเดลโดยมิชอบ
- Data poisoning ซึ่งเป็นการปรับแต่งชุดข้อมูลที่ใช้ในการเรียนรู้ของโมเดลด้วยข้อมูลที่ไม่ถูกต้อง เพื่อทำให้โมเดลสร้างผลลัพธ์ที่ผิดพลาด หรือทำให้เกิดช่องโหว่ที่สามารถถูกโจมตีได้ อันจะทำให้ประสิทธิภาพของระบบ AI ลดลง
- Adversarial attack ซึ่งเป็นการป้อนข้อมูลที่มีลักษณะคล้ายคลึงกับข้อมูลปกติ แต่ส่งผลให้โมเดลทำงานผิดพลาดระหว่างการใช้งาน

ทั้งนี้ ผู้ให้บริการทางการเงินควรพิจารณาความเสี่ยงด้านอื่น ๆ จากการนำระบบ AI มาใช้งาน ซึ่งขึ้นอยู่กับลักษณะการนำไปใช้งาน (use case) เช่น ความเสี่ยงด้านเครดิต ความเสี่ยงด้านปฏิบัติการ

4.3 หลักการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI

ผู้ให้บริการทางการเงินควรบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ให้เหมาะสมกับลักษณะการใช้งาน โดยสอดคล้องกับหลักการ ดังนี้

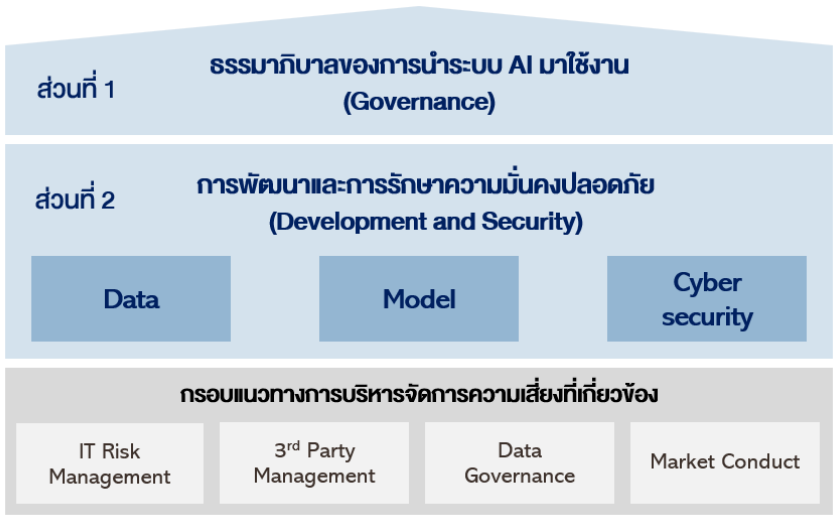
(1) ผู้ให้บริการทางการเงินมีความรับผิดชอบต่อการทำงานของระบบ AI รวมถึงการตัดสินใจที่อ้างอิงจากผลลัพธ์ของระบบ AI

(2) ผู้ให้บริการทางการเงินมีการกำกับดูแลการใช้งานระบบ AI ให้เป็นไปตามหลักการใช้งานระบบ AI อย่างรับผิดชอบ (responsible AI) ที่ได้รับการยอมรับโดยทั่วไป เช่น หลักการใช้งานระบบ AI ที่มีความเป็นธรรม (fairness) มีจริยธรรม (ethics) มีความรับผิดชอบ (accountability) และมีความโปร่งใส (transparency) (หลักการ FEAT)

(3) ผู้ให้บริการทางการเงินมีการบริหารจัดการความเสี่ยงของระบบ AI ครอบคลุมกระบวนการพัฒนาและการใช้งานระบบ AI (AI lifecycle) อย่างรัดกุมเหมาะสม และระบบ AI มีความมั่นคงปลอดภัย ถูกต้อง และเชื่อถือได้ โดยสามารถควบคุมความเสี่ยงและอธิบายที่มาของผลลัพธ์จากระบบ AI ได้

(4) ผู้ให้บริการทางการเงินมีแนวทางการดูแลและคุ้มครองลูกค้าที่โปร่งใสและเป็นธรรม

4.4 แนวทางบริหารจัดการความเสี่ยงของการใช้งานระบบ AI



ภาพรวมแนวทางบริหารจัดการความเสี่ยงของการใช้งานระบบ AI

<u>ส่วนที่ 1</u> ธรรมาภิบาลของการนำระบบ AI มาใช้งาน (governance)

ผลลัพธ์ที่คาดหวัง : ผู้ให้บริการทางการเงินกำหนดความรับผิดชอบในการใช้งานระบบ AI อย่างชัดเจน และมีการกำกับดูแลระบบ AI สอดคล้องตามหลักการ responsible AI ที่ได้รับการยอมรับโดยทั่วไป เช่น หลักการ FEAT รวมถึงมีกรอบแนวทางการบริหารจัดการความเสี่ยงจากการพัฒนาและการใช้งานระบบ AI อย่างรัดกุมเหมาะสม โดยผู้ให้บริการทางการเงินควรดำเนินการ ดังนี้

(1) การกำหนดบทบาทหน้าที่และขอบเขตความรับผิดชอบของคณะกรรมการและผู้บริหารระดับสูงอย่างชัดเจน โดยคณะกรรมการและผู้บริหารระดับสูงมีความรับผิดชอบต่อการตัดสินใจและการดำเนินการต่าง ๆ ของระบบ AI รวมทั้งมีความเข้าใจและตระหนักถึงความเสี่ยงจากการพัฒนาและการใช้งานระบบ AI ตลอดจนมีบทบาทหน้าที่และความรับผิดชอบในการกำกับดูแลความเสี่ยงให้อยู่ในระดับที่องค์กรยอมรับได้ (risk appetite) โดยครอบคลุมการดำเนินการ ดังนี้

(1.1) ดูแลให้มีนโยบายการใช้งานระบบ AI และติดตามการพัฒนาและการใช้งานระบบ AI ให้เป็นไปตามนโยบายขององค์กร เพื่อให้มั่นใจได้ว่าการใช้งานระบบ AI สอดคล้องกับวัตถุประสงค์ของการนำระบบ AI มาใช้งาน และเป็นไปตามนโยบายขององค์กร

(1.2) ดูแลให้มีการกำหนดบทบาทหน้าที่ในการกำกับดูแลและการบริหารความเสี่ยงในการพัฒนาและการใช้งานระบบ AI อย่างชัดเจน ครอบคลุมหน่วยงานที่มีหน้าที่ความรับผิดชอบทั้ง 3 ระดับตามหลักการ three lines of defence

(1.3) ดูแลให้มีการพัฒนาความรู้และทักษะของบุคลากรในองค์กร เพื่อให้บุคลากรสามารถใช้งานระบบ AI ได้อย่างรัดกุมและปลอดภัย โดยไม่พึ่งพาการใช้งานและผลลัพธ์ของระบบ AI มากเกินไปจนก่อให้เกิดความเสี่ยงต่อการดำเนินธุรกิจและการให้บริการแก่ลูกค้า ตลอดจนเพื่อให้เกิดความตระหนักรู้ให้เท่าทันภัยหรือรูปแบบความเสี่ยงที่อาจเกิดขึ้น

(2) การกำหนดนโยบายและกรอบแนวทางการใช้งานและการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ขององค์กรอย่างชัดเจน สอดคล้องกับเป้าหมายขององค์กร และเป็นไปตามหลักการ responsible AI ที่ได้รับการยอมรับโดยทั่วไป โดยนโยบายมีความสอดคล้องกับกลยุทธ์การใช้งานระบบ AI ขององค์กร บทบาทหน้าที่ของผู้ที่เกี่ยวข้อง ตลอดจนแนวทางการกำกับดูแลการพัฒนาและการใช้งานระบบ AI ขององค์กร รวมทั้งจัดให้มีการทบทวนนโยบายดังกล่าวเป็นประจำ เพื่อให้เท่าทันกับพัฒนาการของเทคโนโลยี รูปแบบลักษณะการใช้งาน และความเสี่ยงที่เปลี่ยนแปลงไป

(3) การบริหารจัดการความเสี่ยงจากการใช้งานระบบ AI โดยครอบคลุมการดำเนินการ ดังนี้

(3.1) ระบุความเสี่ยงจากการใช้งานระบบ AI และกำหนดเป้าหมายของ risk appetite ขององค์กรอย่างชัดเจน รวมทั้งจัดให้มีการประเมินและติดตามความเสี่ยงของการใช้งานระบบ AI อย่างต่อเนื่อง เพื่อให้มีการควบคุมความเสี่ยงเหมาะสมกับลักษณะการใช้งาน ภายใต้กรอบนโยบายการบริหารจัดการความเสี่ยงที่ยอมรับได้ขององค์กร

(3.2) พิจารณาความเสี่ยงและผลกระทบของการใช้งานระบบ AI ที่มีต่อการดำเนินธุรกิจและการให้บริการแก่ลูกค้า โดยเฉพาะกรณีที่มีการนำระบบ AI มาใช้กับงานหลักที่เกี่ยวข้องกับการตัดสินใจเชิงกลยุทธ์[1] หรืองานที่ส่งผลกระทบโดยตรงต่อลูกค้า โดยในการนำระบบ AI มาใช้ในลักษณะดังกล่าว อาจพิจารณาให้มนุษย์มีส่วนร่วมในการตัดสินใจ (human in the loop[2]) หรือกำกับดูแล (human over the loop[3]) ตามความเหมาะสม แล้วแต่กรณี โดยพิจารณาระดับการมีส่วนร่วมหรือการกำกับดูแลตามความเสี่ยงและผลกระทบของการใช้งาน

---

[1] งานหลักที่เกี่ยวข้องกับการตัดสินใจเชิงกลยุทธ์ อ้างอิงตามประกาศธนาคารแห่งประเทศไทยว่าด้วยหลักเกณฑ์การใช้บริการจากพันธมิตรทางธุรกิจ (business partner) ของสถาบันการเงิน และประกาศธนาคารแห่งประเทศไทยว่าด้วยหลักเกณฑ์การกำกับดูแลการใช้บริการจากผู้ให้บริการสนับสนุนการประกอบธุรกิจ (business facilitator) ของสถาบันการเงินเฉพาะกิจ เช่น งานอนุมัติสินเชื่อ งานอนุมัติเปิดบัญชี งานอนุมัติการฝาก ถอน หรือโอนเงิน

[2] human in the loop หมายความว่า การใช้งานระบบ AI ที่มนุษย์มีส่วนร่วมในการควบคุมการทำงานหรือตัดสินใจทั้งหมด โดยมีระบบ AI ทำหน้าที่ให้คำแนะนำหรือข้อมูล แต่ไม่สามารถทำงานหรือตัดสินใจในการดำเนินการใด ๆ ได้โดยปราศจากมนุษย์

[3] human over the loop หมายความว่า การใช้งานระบบ AI ที่ระบบ AI สามารถทำงานหรือตัดสินใจได้โดยอัตโนมัติ แต่ยังคงจำเป็นต้องมีมนุษย์ในการกำกับดูแล และมนุษย์สามารถเข้าควบคุมหรือระงับการทำงานได้เมื่อพบความผิดพลาดหรืออาจก่อให้เกิดผลกระทบเชิงลบ

(3.3) ในการนำระบบ AI มาใช้สนทนาหรือสื่อสารกับลูกค้าแทนมนุษย์ มีแนวทางการแจ้งเกี่ยวกับการให้บริการด้วยระบบ AI ให้ลูกค้าทราบ รวมทั้งอาจมีทางเลือกสำหรับลูกค้าในการติดต่อกับเจ้าหน้าที่ในกรณีที่ลูกค้าไม่ประสงค์จะสนทนาหรือสื่อสารกับระบบ AI

<u>ส่วนที่ 2</u> การพัฒนาและการรักษาความมั่นคงปลอดภัยของการใช้งานระบบ AI (development and security)

ผลลัพธ์ที่คาดหวัง : ผู้ให้บริการทางการเงินมีแนวทางควบคุมความเสี่ยงของการใช้งานระบบ AI ครอบคลุมด้านข้อมูลที่ระบบ AI ใช้ในการเรียนรู้ ด้านการพัฒนาโมเดล และด้านภัยคุกคามทางไซเบอร์ เพื่อให้การใช้งานระบบ AI อยู่ภายใต้กรอบความถูกต้อง เชื่อถือได้ และโปร่งใส ตลอดจนมีความมั่นคงปลอดภัยจากภัยคุกคามไซเบอร์รูปแบบใหม่ที่อาจเกิดขึ้น โดยผู้ให้บริการทางการเงินควรดำเนินการดังนี้

(1) การควบคุมความเสี่ยงด้านข้อมูล โดยครอบคลุมการดำเนินการ ดังนี้

(1.1) มีแนวทางควบคุมและประเมินคุณภาพข้อมูลที่ใช้ในการเรียนรู้ของโมเดล โดยจัดทำเกณฑ์การประเมินคุณภาพข้อมูลอย่างชัดเจน ครอบคลุมมิติต่าง ๆ ทั้งความเป็นปัจจุบัน ความถูกต้อง ตลอดจนปริมาณและความหลากหลายของข้อมูล เพื่อลดความเสี่ยงที่โมเดลอาจสร้างผลลัพธ์ที่ไม่น่าเชื่อถือ เชิงลบ โน้มเอียง หรือเป็นเท็จ

(1.2) จัดให้มีขั้นตอนการตรวจสอบและประเมินคุณภาพข้อมูล ก่อนนำมาใช้ในกระบวนการเรียนรู้ของโมเดล เพื่อให้สามารถปรับปรุงแก้ไขข้อมูลให้มีคุณภาพตามเกณฑ์การประเมินคุณภาพข้อมูล

(1.3) มีแนวทางป้องกันข้อมูลถูกเปิดเผยแก่ผู้ที่ไม่เกี่ยวข้องหรือรั่วไหลจากระบบ AI อย่างน้อยครอบคลุม

- การจำกัดการให้ระบบ AI เข้าถึงข้อมูลขององค์กร (access control) เพื่อป้องกันการที่ระบบ AI เข้าถึงข้อมูลสำคัญโดยไม่จำเป็น ตลอดจนมีการควบคุมสิทธิของบุคลากรในการเข้าถึงข้อมูลที่ใช้ในการเรียนรู้และข้อมูลโครงสร้างการทำงานของโมเดลในระบบ AI
- การจัดให้มีแนวทางหรือวิธีการปกป้องข้อมูลสำคัญ ข้อมูลส่วนบุคคล หรือข้อมูลอ่อนไหว ไม่ให้ถูกเปิดเผยแก่ผู้ที่ไม่เกี่ยวข้องหรือรั่วไหลออกไปภายนอก เช่น การปกปิดข้อมูล (data masking) การแฮชข้อมูล (data hashing) การควบคุมข้อมูลนำเข้า (input sanitization)

- การจัดให้มีแนวทางในการแยกข้อมูลองค์กรออกจากข้อมูลสาธารณะ ในกรณีที่ใช้ระบบ AI ของบุคคลภายนอก เช่น การจำกัดหรือกำหนดขอบเขตของข้อมูล (data boundary) เพื่อแยกข้อมูลองค์กรที่ใช้ในการประมวลผลของ Generative AI[4]

(2) การควบคุมความเสี่ยงด้านการพัฒนาโมเดล โดยครอบคลุมการดำเนินการ ดังนี้

(2.1) มีเกณฑ์การควบคุมและประเมินประสิทธิภาพของโมเดล โดยจัดทำตัวชี้วัดประสิทธิภาพ (evaluation metrics) ของผลลัพธ์อย่างชัดเจน และครอบคลุมด้านความถูกต้องแม่นยำและด้านความน่าเชื่อถือ

(2.2) ทดสอบและติดตามประสิทธิภาพและความน่าเชื่อถือของผลลัพธ์ของโมเดลทั้งก่อนและหลังการใช้งานอย่างต่อเนื่อง เพื่อให้มั่นใจว่าโมเดลมีความสม่ำเสมอในการสร้างผลลัพธ์ โดยอาจทดสอบด้วยข้อมูลหลากหลายรูปแบบ เช่น ข้อมูลชุดใหม่ (unseen data) ข้อมูลที่มีลักษณะคลาดเคลื่อนจากข้อมูลที่ใช้ในการเรียนรู้ (noise) ข้อมูลกรณีพิเศษที่อาจมีโอกาสพบเจอได้ยาก (edge case) รวมทั้งอาจพิจารณาให้มีกระบวนการนำผลลัพธ์หรือการตัดสินใจของโมเดลกลับมาใช้ (feedback loop) เพื่อนำผลลัพธ์มาปรับปรุงและพัฒนาโมเดลให้มีความถูกต้องแม่นยำและน่าเชื่อถือต่อเนื่อง

(2.3) ในการใช้ Generative AI มีแนวทางลดความเสี่ยงจากการที่ระบบให้ข้อมูลเท็จ (hallucination) เช่น การใช้เทคนิค retrieval-augmented generation[5] เพื่อช่วยให้โมเดลมีแหล่งข้อมูลสำหรับอ้างอิง หรือการสร้างและปรับแต่งข้อความหรือคำสั่งเพื่อช่วยกำหนดรูปแบบหรือจำกัดขอบเขตของผลลัพธ์ของโมเดล (prompt engineering)

(2.4) มีแนวทางให้ผู้พัฒนา ผู้ใช้งาน และผู้ที่เกี่ยวข้อง สามารถเข้าใจและอธิบายผลลัพธ์ของระบบ AI เช่น การจัดทำเอกสารเพื่ออธิบายข้อมูลขาเข้า (input) ผลลัพธ์ขาออก (output) และพารามิเตอร์ที่โมเดลใช้ หรือการให้โมเดลประเภท Generative AI แสดงกระบวนการคิด (reasoning) หรือระบุแหล่งอ้างอิงของเนื้อหา (referencing) ควบคู่กับการแสดงผลลัพธ์

---

[4] Generative AI หมายความว่า ระบบ AI ที่มีความสามารถในการสร้างเนื้อหาใหม่ในหลากหลายรูปแบบ เช่น ข้อความ ภาพ วีดีโอ ซอร์สโค้ด หรือรูปแบบอื่น ตามข้อความหรือคำสั่ง (prompt) ที่มนุษย์เป็นผู้กำหนด

[5] Retrieval-augmented generation (RAG) หมายความว่า เทคนิคปรับปรุงผลลัพธ์ของโมเดลประเภท large language model ด้วยการให้โมเดลอ้างอิงฐานความรู้ที่เชื่อถือได้ซึ่งอยู่นอกเหนือจากแหล่งข้อมูลที่ใช้ในการเรียนรู้ก่อนที่จะแสดงผลลัพธ์ เพื่อควบคุมผลลัพธ์และสามารถเข้าใจวิธีการที่โมเดลสร้างผลลัพธ์อย่างชัดเจนมากขึ้น

(3) การควบคุมความเสี่ยงด้านภัยคุกคามทางไซเบอร์ โดยครอบคลุมการดำเนินการ ดังนี้

(3.1) จัดการป้องกันไม่ให้ระบบ AI แสดงข้อมูลสำคัญขององค์กรที่ไม่พึงเปิดเผยหรือข้อมูลเชิงลบ เช่น การคัดกรองเนื้อหา (content filtering) กับระบบ AI ทั้งการส่งคำสั่งขาเข้า (prompt filtering) และแสดงผลลัพธ์ขาออก (response filtering)

(3.2) จัดการทดสอบรูปแบบการโจมตีต่าง ๆ อย่างต่อเนื่อง เพื่อให้มั่นใจได้ว่าระบบ AI ทำงานถูกต้องตามวัตถุประสงค์ของการใช้งาน รวมทั้งไม่มีช่องโหว่ที่กระทบต่อความปลอดภัยของโมเดลและระบบเทคโนโลยีสารสนเทศ (IT) ขององค์กร เช่น ทดสอบเพื่อป้องกัน data poisoning หรือ adversarial attack

(3.3) ติดตามและประเมินภัยคุกคามใหม่ ๆ อย่างต่อเนื่อง พร้อมทั้งปรับปรุงแนวทางป้องกันและตรวจจับให้เท่าทันกับภัยคุกคามที่อาจเกิดขึ้นกับระบบ AI โดยสามารถอ้างอิงจากมาตรฐานสากล เช่น OWASP Machine Learning Security Top 10, OWASP Top 10 for Large Language Model Applications, MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems)

คำถาม – คำตอบแนบท้ายแนวนโยบายธนาคารแห่งประเทศไทย

เรื่อง การบริหารจัดการความเสี่ยงของการใช้งานระบบปัญญาประดิษฐ์

ลำดับ	คำถาม	คำตอบ
1	ขอทราบตัวอย่างของระบบ AI ที่ไม่ครอบคลุมตามนิยามในแนวนโยบายฉบับนี้	ระบบ AI ในแนวนโยบายฉบับนี้ ไม่ครอบคลุมถึงระบบอัตโนมัติที่มนุษย์เป็นผู้กำหนดขั้นตอนการทำงานของระบบ เช่น Robotic Process Automation (RPA) หรือระบบที่ทำงานอัตโนมัติเมื่อเข้าเงื่อนไขที่กำหนดล่วงหน้า (condition matching) หรือสคริปต์อัตโนมัติ (automated script) เป็นต้น
2	ขอทราบรายละเอียดเกี่ยวกับแนวทางการประเมินคุณภาพข้อมูลที่ใช้ในการเรียนรู้ของโมเดล	ผู้ให้บริการทางการเงินสามารถนำแนวทางการบริหารจัดการคุณภาพข้อมูลตามแนวนโยบาย ธปท. ว่าด้วยการกำกับดูแลข้อมูล (Data Governance) มาอ้างอิงและประยุกต์ใช้ได้ตามความเหมาะสม
3	ขอทราบตัวอย่างรูปแบบข้อมูลสำหรับการทดสอบและติดตามประสิทธิภาพและความน่าเชื่อถือของผลลัพธ์ของโมเดล	ผู้ให้บริการทางการเงินควรทดสอบความน่าเชื่อถือของระบบ AI ด้วยข้อมูลหลากหลายรูปแบบ เช่น ข้อมูลชุดใหม่ที่ไม่เคยปรากฏในข้อมูลที่ใช้ในการเรียนรู้มาก่อน (unseen data) ข้อมูลกรณีพิเศษที่อาจมีโอกาสพบเจอได้ยากหรือมีค่าสูงต่ำผิดปกติ (edge case) หรือข้อมูลที่มีลักษณะคล้ายคลึงข้อมูลปกติแต่อาจทำให้ระบบทำงานผิดพลาด (adversarial data)
4	แนวนโยบายฉบับนี้มีข้อกำหนดสำหรับผู้ให้บริการทางการเงินในการบริหารจัดการความเสี่ยงทางกฎหมายจากการใช้งานระบบ AI อย่างไร	แนวนโยบายฉบับนี้กำหนดหลักการและแนวทางในการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI สำหรับผู้ให้บริการทางการเงินสามารถนำไปประยุกต์ใช้ตามลักษณะการใช้งาน เพื่อให้การใช้งานระบบ AI มีความเหมาะสมและสอดคล้องกับหลักการที่ดีที่ได้รับการยอมรับในระดับสากล อย่างไรก็ดี ผู้ให้บริการทางการเงินต้องปฏิบัติตามกฎหมายที่เกี่ยวข้องอย่างเคร่งครัด เช่น กฎหมายคุ้มครองข้อมูลส่วนบุคคล กฎหมายทรัพย์สินทางปัญญา

5	หากผู้ให้บริการทางการเงินไม่ได้เป็นผู้พัฒนาระบบ AI ขึ้นเอง โดยนำระบบ AI แบบสำเร็จรูปมาใช้งานเท่านั้น เช่น การนำ Generative AI chatbot มาใช้งานภายในองค์กร ผู้ให้บริการทางการเงินต้องปฏิบัติตามแนวนโยบายฉบับนี้หรือไม่	แนวนโยบายฉบับนี้กำหนดเป็นหลักการและแนวทางในการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI โดยผู้ให้บริการทางการเงินสามารถพิจารณานำเนื้อหาในส่วนที่เกี่ยวข้องมาใช้ในการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ของตนได้ ไม่ว่าจะเป็นระบบ AI ที่ผู้ให้บริการทางการเงินพัฒนาขึ้นเอง หรือระบบ AI ของบุคคลภายนอกที่มีการนำมาใช้งาน