(ร่าง) แนวนโยบายธนาคารแห่งประเทศไทย
เรื่อง การบริหารจัดการความเสี่ยงของการใช้งานระบบปัญญาประดิษฐ์
(Guiding Principles for Artificial Intelligence Risk Management)

ธนาคารแห่งประเทศไทย

จัดทำโดย

ฝ่ายกำกับและตรวจสอบความเสี่ยงด้านเทคโนโลยีสารสนเทศ

สายกำกับระบบการชำระเงินและคุ้มครองผู้ใช้บริการทางการเงิน

ธนาคารแห่งประเทศไทย

โทรศัพท์ 0-2283-6469, 0-2283-6059

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

สารบัญ

แนวนโยบายธนาคารแห่งประเทศไทย
เรื่อง การบริหารจัดการความเสี่ยงของการใช้งานระบบปัญญาประดิษฐ์
(Guiding Principles for Artificial Intelligence Risk Management)

1. เหตุผลในการออกแนวนโยบาย

ในปัจจุบันผู้ให้บริการทางการเงินใช้ระบบปัญญาประดิษฐ์ (Artificial Intelligence: AI) เป็นเครื่องมือช่วยเพิ่มประสิทธิภาพในการดำเนินธุรกิจและการให้บริการแก่ลูกค้า โดยนำระบบ AI มาใช้สนับสนุนงานสำคัญหรืองานให้บริการลูกค้ามากขึ้น อย่างไรก็ตาม การนำระบบ AI มาใช้ อาจทำให้รูปแบบความเสี่ยงเดิมของผู้ให้บริการทางการเงินถูกขยายหรือปรับเปลี่ยนไป เนื่องจากระบบ AI สามารถเรียนรู้และต่อยอดจากข้อมูลเพื่อสร้างผลลัพธ์ที่นำมาประกอบการตัดสินใจหรือทำงานแทนมนุษย์ ดังนั้นผู้ให้บริการทางการเงินจำเป็นต้องควบคุมและจัดการความเสี่ยง เพื่อให้มั่นใจว่าระบบ AI สามารถให้ผลลัพธ์ที่ถูกต้อง น่าเชื่อถือ โปร่งใสและอธิบายได้

ธนาคารแห่งประเทศไทย (ธปท.) สนับสนุนให้ผู้ให้บริการทางการเงินนำระบบ AI มาใช้ในการเพิ่มประสิทธิภาพการปฏิบัติงานและการดำเนินธุรกิจ ตลอดจนการพัฒนานวัตกรรมที่สามารถตอบสนองความต้องการของลูกค้าได้ดียิ่งขึ้น ในขณะเดียวกันผู้ให้บริการทางการเงินต้องมีการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ที่รัดกุม เพื่อไม่ให้มีความเสี่ยงต่อการดำเนินธุรกิจ รวมทั้งผู้ใช้บริการได้รับการดูแลและคุ้มครองอย่างเหมาะสม

ธปท. จึงออกแนวนโยบายเรื่องการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ฉบับนี้ เพื่อให้ผู้ให้บริการทางการเงินนำไปใช้อ้างอิงเป็นแนวทางในการบริหารความเสี่ยงอย่างเหมาะสม และสอดคล้องกับหลักการที่ดีที่ได้รับการยอมรับในระดับสากล

2. ผลลัพธ์ที่คาดหวัง

ผู้ให้บริการทางการเงินมีแนวทางที่สามารถนำไปประยุกต์ใช้เพื่อบริหารจัดการความเสี่ยงของการใช้งานระบบ AI เหมาะสมตามลักษณะการนำไปใช้งาน

โดยแนวนโยบายฉบับนี้เป็นส่วนเพิ่มเติมจากแนวทางบริหารจัดการความเสี่ยงเดิมที่ผู้ให้บริการทางการเงินพึงปฏิบัติอยู่ ได้แก่ แนวทางบริหารจัดการความเสี่ยงด้านเทคโนโลยีสารสนเทศ (Information Technology Risk Management) แนวทางบริหารจัดการความเสี่ยงจากบุคคลภายนอก (Third Party Risk Management) แนวทางการกำกับดูแลข้อมูล (Data Governance) และแนวทางบริหารจัดการด้านการให้บริการแก่ลูกค้าอย่างเป็นธรรม (Market conduct)

3. ขอบเขตการใช้แนวนโยบาย

แนวนโยบายฉบับนี้ใช้กับสถาบันการเงินและสถาบันการเงินเฉพาะกิจตามกฎหมายว่าด้วยธุรกิจสถาบันการเงิน ผู้ประกอบธุรกิจระบบและบริการการชำระเงินภายใต้การกำกับตามกฎหมายว่าด้วยระบบการชำระเงิน

4. เนื้อหา

4.1 นิยาม

ในแนวนโยบายฉบับนี้

“ระบบปัญญาประดิษฐ์หรือระบบ AI” หมายถึง ระบบที่เลียนแบบปัญญาของมนุษย์สามารถทำงานต่าง ๆ เช่น เรียนรู้ จดจำ ตัดสินใจ ปฏิบัติงาน และสร้างสรรค์เนื้อหาใหม่ได้ ผ่านกลไกการเรียนรู้จากข้อมูลและการสร้างเงื่อนไขอัตโนมัติ โดยครอบคลุมถึงระบบที่มีส่วนประกอบของโมเดลประเภท Machine Learning, Deep Learning ตลอดจน Generative AI เช่น Large Language Model (LLM) และระบบขั้นสูง เช่น Agentic AI ทั้งนี้ไม่ครอบคลุมระบบอัตโนมัติที่มนุษย์เป็นผู้กำหนดขั้นตอนการทำงาน เช่น Robotic Process Automation (RPA) หรือระบบที่ทำงานอัตโนมัติเมื่อเข้าเงื่อนไขที่กำหนดล่วงหน้า (condition matching) เป็นต้น

“ผู้ให้บริการทางการเงิน” หมายถึง สถาบันการเงินและสถาบันการเงินเฉพาะกิจตามกฎหมายว่าด้วยธุรกิจสถาบันการเงิน และผู้ประกอบธุรกิจระบบและบริการการชำระเงินภายใต้การกำกับตามกฎหมายว่าด้วยระบบการชำระเงิน

4.2 ความเสี่ยงจากการใช้งานระบบ AI

ความเสี่ยงของการใช้งานระบบ AI สามารถเกิดขึ้นได้ในกระบวนการเรียนรู้ ตั้งแต่ขั้นตอนการเตรียมข้อมูลและการพัฒนาโมเดล ตลอดจนการนำระบบ AI ไปใช้งาน โดยจำแนกความเสี่ยงออกเป็น 3 ด้าน ดังนี้

1) ความเสี่ยงด้านข้อมูล หมายถึง ความเสี่ยงที่ข้อมูลสำหรับการเรียนรู้ของโมเดลไม่มีคุณภาพ เช่น ไม่เป็นปัจจุบัน ไม่ถูกต้อง ไม่หลากหลายเพียงพอ ส่งผลให้โมเดลสร้างผลลัพธ์ที่ไม่น่าเชื่อถือ เชิงลบ โน้มเอียง หรือเป็นเท็จ นอกจากนี้การรักษาความลับและความมั่นคงปลอดภัยของข้อมูลในขั้นตอนการเรียนรู้หรือการใช้งานระบบ AI ที่ไม่รัดกุมเพียงพอ อาจส่งผลให้ข้อมูลสำคัญขององค์กรหรือข้อมูลส่วนบุคคลของลูกค้ารั่วไหลไปภายนอกได้
2) ความเสี่ยงด้านการพัฒนาโมเดล หมายถึง ความเสี่ยงที่โมเดลอาจเรียนรู้และสร้างเงื่อนไขที่ไม่สามารถอธิบายที่มาของผลลัพธ์ได้ รวมทั้งอาจสร้างผลลัพธ์ที่ไม่ถูกต้อง

ไม่น่าเชื่อถือ เช่น ผลลัพธ์ที่มีความโน้มเอียง ไม่แน่นอน ไม่แม่นยำ หรือให้ข้อมูลที่เป็นเท็จ (hallucination) จนส่งผลให้เกิดความเสี่ยงต่อการดำเนินธุรกิจ ความเสี่ยงด้านการปฏิบัติงาน และความเสี่ยงที่การให้บริการลูกค้าอาจไม่สอดคล้องตามแนวทางการให้บริการลูกค้าอย่างเป็นธรรม

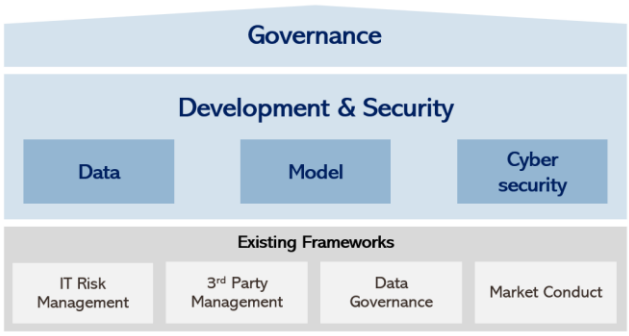
3) ความเสี่ยงที่เกิดจากภัยคุกคามทางไซเบอร์ หมายถึง ความเสี่ยงจากการโจมตีรูปแบบใหม่ที่เจาะจงกับระบบ AI เพื่อเข้าถึงและให้ระบบแสดงข้อมูลสำคัญโดยไม่ได้รับอนุญาต หรือมุ่งประสงค์ให้ระบบประมวลผลผิดพลาดหรือบิดเบือนผลลัพธ์ รวมทั้งเพื่อค้นหาช่องโหว่ด้านความปลอดภัย โดยใช้รูปแบบการโจมตีต่าง ๆ เช่น การโจมตีระบบผ่านคำสั่ง (prompt injection) การเข้าถึงโครงสร้างการทำงานของโมเดล (model inversion) การแทรกหรือแก้ไขข้อมูลสำหรับเรียนรู้ระหว่างพัฒนาโมเดล (data poisoning) การป้อนข้อมูลปรับแต่งที่มีวัตถุประสงค์ให้โมเดล AI ทำงานผิดพลาดระหว่างใช้งาน (adversarial attack) เป็นต้น

4.3 หลักการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI

ผู้ให้บริการทางการเงินควรมีการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI ให้เหมาะสมกับลักษณะการนำไปใช้งาน โดยมีหลักการดังนี้

1. รับผิดชอบต่อการดำเนินการและผลจากการตัดสินใจของระบบ AI โดยนำระบบ AI มาใช้เพิ่มประสิทธิภาพการดำเนินธุรกิจและการให้บริการลูกค้าภายใต้กรอบการกำกับดูแลที่เหมาะสม สอดคล้องกับหลักการใช้งาน AI อย่างรับผิดชอบ มีจริยธรรม โปร่งใส เป็นธรรม (FEAT: fairness, ethics, accountability, transparency)
2. มีการบริหารจัดการความเสี่ยงของระบบ AI ครอบคลุมกระบวนการพัฒนาและการใช้งานระบบ AI (AI lifecycle) อย่างรัดกุมเหมาะสม ระบบมีความมั่นคงปลอดภัย ถูกต้อง เชื่อถือได้ ซึ่งผู้ให้บริการทางการเงินสามารถควบคุมความเสี่ยงและอธิบายที่มาของผลลัพธ์จากระบบ AI ได้
3. มีแนวทางการดูแลและคุ้มครองผู้ใช้บริการอย่างโปร่งใสและเป็นธรรม

4.4 แนวนโยบายการบริหารจัดการความเสี่ยงของการใช้งานระบบ AI



ภาพรวมแนวนโยบายการบริหารจัดการความเสี่ยงของการใช้งานระบบ

แนวนโยบายฉบับนี้ แบ่งออกเป็น 2 ส่วน ดังนี้

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

วัตถุประสงค์: ผู้ให้บริการทางการเงินกำหนดความรับผิดชอบอย่างชัดเจน และมีการกำกับดูแลระบบ AI สอดคล้องตามหลัก FEAT รวมถึงมีการบริหารจัดการความเสี่ยงจากการพัฒนาและใช้งานระบบ AI อย่างรัดกุมเหมาะสม

ผู้ให้บริการทางการเงิน ควรดำเนินการดังนี้

1) **กำหนดบทบาทหน้าที่และขอบเขตความรับผิดชอบของคณะกรรมการและผู้บริหารระดับสูงอย่างชัดเจน**

คณะกรรมการและผู้บริหารระดับสูงต้องเข้าใจและตระหนักถึงความเสี่ยงจากการพัฒนาและใช้งานระบบ AI รวมทั้งมีบทบาทหน้าที่และความรับผิดชอบในการกำกับดูแลความเสี่ยงให้อยู่ในระดับที่องค์กรยอมรับได้ โดยครอบคลุมการดำเนินการและการดูแล ดังนี้

(1.1) ดูแลให้มีนโยบายการใช้งานระบบ AI และติดตามการพัฒนาและใช้งานระบบ AI ให้เป็นไปตามนโยบายที่กำหนด โดยไม่ก่อให้เกิดความเสี่ยงต่อการดำเนินธุรกิจและการให้บริการลูกค้า

(1.2) ดูแลให้มีการกำหนดบทบาทหน้าที่ในการกำกับดูแล และบริหารความเสี่ยงในการพัฒนาและใช้งานระบบ AI อย่างชัดเจน ทั้งในด้านความรับผิดชอบ มีจริยธรรม โปร่งใส เป็นธรรม และด้านความถูกต้องเชื่อถือได้รวมถึงความมั่นคงปลอดภัยของระบบ AI โดยครอบคลุมบทบาทหน้าที่ทั้ง three lines of defence

(1.3) ดูแลให้เกิดการสร้างความตระหนักรู้ให้ทุกคนในองค์กรใช้งานระบบ AI อย่างระมัดระวังสอดคล้องตามนโยบายที่กำหนด เพื่อให้ใช้งานระบบ AI ได้อย่างเหมาะสม ไม่ก่อให้เกิดความเสี่ยงต่อการการดำเนินงาน รวมถึงรู้เท่าทันรูปแบบความเสี่ยงที่อาจเกิดขึ้นใหม่ ๆ

2) กำหนดนโยบายและแนวทางการใช้งานระบบ AI ขององค์กรอย่างชัดเจน สอดคล้องกับเป้าหมายองค์กร หลักการ FEAT ระเบียบและกฎหมายที่เกี่ยวข้อง โดยนโยบายควรระบุทิศทางกลยุทธ์เกี่ยวกับการนำ AI มาใช้ บทบาทหน้าที่ของผู้ที่เกี่ยวข้อง และแนวทางการกำกับดูแลการพัฒนาและใช้งานระบบ AI รวมทั้งจัดให้มีการทบทวนนโยบายดังกล่าวเป็นประจำ เพื่อให้เท่าทันกับพัฒนาการของเทคโนโลยี รูปแบบลักษณะการใช้งานหรือความเสี่ยงที่เปลี่ยนแปลงไป

3) มีการบริหารจัดการความเสี่ยงจากการใช้งานระบบ AI ครอบคลุม

(3.1) ระบุความเสี่ยงจากการใช้งานระบบ AI และกำหนดเป้าหมาย risk appetite ขององค์กรอย่างชัดเจน รวมทั้งจัดให้มีการประเมินและติดตามความเสี่ยงของการใช้งานระบบ AI อย่างต่อเนื่อง เพื่อให้มีการควบคุมความเสี่ยงเหมาะสมกับลักษณะการใช้งาน ภายใต้กรอบนโยบายการบริหารจัดการความเสี่ยงที่ยอมรับได้

(3.2) กรณีที่มีการนำระบบ AI มาใช้กับงานที่เกี่ยวข้องกับการตัดสินใจเชิงกลยุทธ์ (strategic function[1]) หรืองานที่โต้ตอบกับลูกค้า (customer interaction) ผู้ให้บริการทางการเงินควรพิจารณาให้มนุษย์มีส่วนร่วมในการตัดสินใจ โดยอาจเลือกใช้วิธี human in the loop[2] หรือ human over the loop[3] ขึ้นกับการพิจารณาความเสี่ยงและผลกระทบของการใช้งานระบบ AI ที่มีต่อการดำเนินธุรกิจและลูกค้า

---

[1] strategic function: อ้างอิงตามประกาศธนาคารแห่งประเทศไทยที่ สนส. 16/2563 เรื่อง หลักเกณฑ์การใช้บริการจากพันธมิตรทางธุรกิจ ตัวอย่างงาน strategic function เช่น งานอนุมัติสินเชื่อ งานอนุมัติเปิดบัญชี งานอนุมัติฝากถอนโอน เป็นต้น

[2] human in the loop: การใช้งาน AI ที่มนุษย์มีส่วนร่วมในการควบคุมการทำงานหรือตัดสินใจทั้งหมด โดยมี AI ทำหน้าที่ให้คำแนะนำหรือข้อมูล แต่ไม่สามารถทำงานหรือตัดสินใจในการดำเนินการใด ๆ ได้ โดยปราศจากมนุษย์ (อ้างอิงจาก AI Governance Guideline for Executives โดย ETDA)

[3] human over the loop: การใช้งาน AI ที่ AI สามารถทำงานหรือตัดสินใจได้โดยอัตโนมัติ แต่ยังคงจำเป็นต้องมีมนุษย์ในการกำกับดูแล และมนุษย์สามารถเข้าควบคุมหรือระงับการทำงานได้เมื่อพบความผิดพลาดหรืออาจก่อให้เกิดผลกระทบเชิงลบ (อ้างอิงจาก AI Governance Guideline for Executives โดย ETDA)

(3.3) กรณีใช้งานระบบ AI กับงานที่โต้ตอบกับลูกค้า ผู้ให้บริการทางการเงินควรแจ้งและเปิดเผยเกี่ยวกับการให้บริการด้วยระบบ AI ต่อลูกค้าก่อนให้บริการ และมีตัวเลือกให้ลูกค้าสามารถปิด (disable) หรือข้าม (bypass) บริการที่ใช้งานร่วมกับระบบ AI ได้ รวมทั้งจัดให้มีการให้ความรู้เกี่ยวกับวิธีการใช้งานอย่างปลอดภัยแก่ลูกค้า

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

วัตถุประสงค์: ผู้ให้บริการทางการเงินมีแนวทางควบคุมความเสี่ยงข้อมูลที่นำมาใช้ในการเรียนรู้ การพัฒนาโมเดลและการใช้งานระบบ AI ให้มีความถูกต้อง เชื่อถือได้ โปร่งใส และมั่นคงปลอดภัยจากภัยคุกคามไซเบอร์รูปแบบใหม่ที่อาจเกิดขึ้น

ผู้ให้บริการทางการเงินควรดำเนินการดังนี้

1) การควบคุมความเสี่ยงด้านข้อมูล

(1.1) มีแนวทางควบคุมและประเมินคุณภาพข้อมูลที่ใช้ในการเรียนรู้ของโมเดล โดยจัดทำหลักเกณฑ์เพื่อใช้ในการประเมินคุณภาพข้อมูลที่นำมาใช้ในการเรียนรู้ของโมเดลอย่างชัดเจน และครอบคลุมมิติด้านความถูกต้อง ความเป็นปัจจุบัน ปริมาณและความหลากหลายของข้อมูล เพื่อลดความเสี่ยงที่โมเดลอาจสร้างผลลัพธ์ที่ไม่น่าเชื่อถือ เชิงลบ โน้มเอียง หรือเป็นเท็จ

(1.2) ดำเนินการตรวจสอบและประเมินคุณภาพข้อมูล รวมทั้งมีแนวทางแก้ไขข้อมูลให้เป็นไปตามหลักเกณฑ์ที่กำหนด ก่อนนำมาใช้ในกระบวนการเรียนรู้ของโมเดล

(1.3) มีแนวทางป้องกันข้อมูลรั่วไหลจากระบบ AI อย่างน้อยครอบคลุม

- จำกัดสิทธิ์ของระบบ AI ในการเข้าถึงข้อมูล (access control) ขององค์กร เพื่อป้องกันการเข้าถึงข้อมูลสำคัญโดยไม่จำเป็น ตลอดจนควบคุมสิทธิ์การเข้าถึงข้อมูลสำหรับเรียนรู้และโครงสร้างการทำงานของโมเดล

- มีแนวทางปกป้องข้อมูลส่วนบุคคลหรือข้อมูลอ่อนไหว ตั้งแต่การเตรียมข้อมูล การพัฒนาโมเดล และการใช้งานระบบ AI เช่น ทำ data masking, data hashing, input sanitization, ปรับปรุงระบบ Data Leak Prevention (DLP) ให้ครอบคลุมผลลัพธ์ของระบบ AI เพื่อป้องกันการรั่วไหลของข้อมูลดังกล่าว
- มีแนวทางในการแยกข้อมูลองค์กรออกจากข้อมูลสาธารณะ กรณีใช้ระบบ AI ของบุคคลภายนอก เช่น การทำ data boundary เพื่อแยกข้อมูลองค์กรในการใช้งาน Generative AI

2) การควบคุมความเสี่ยงด้านการพัฒนาโมเดล

(2.1) มีเกณฑ์การควบคุมและประเมินประสิทธิภาพของโมเดล โดยจัดทำตัวชี้วัดประสิทธิภาพ (evaluation metrics) อย่างชัดเจนและครอบคลุมด้านความถูกต้องแม่นยำและด้านความน่าเชื่อถือ เหมาะสมกับความเสี่ยงและผลกระทบของลักษณะการนำไปใช้งาน

(2.2) ทดสอบและติดตามประสิทธิภาพและความน่าเชื่อถือของผลลัพธ์ ทั้งก่อนและหลังใช้งานอย่างต่อเนื่อง เพื่อให้มั่นใจว่าโมเดลมีความสม่ำเสมอในการให้ผลลัพธ์ที่ถูกต้องแม่นยำและน่าเชื่อถือ เป็นไปตามเกณฑ์ที่กำหนด โดยอาจทดสอบด้วยข้อมูลหลากหลายรูปแบบ เช่น ข้อมูลชุดใหม่ (unseen data) ข้อมูลที่มีลักษณะคลาดเคลื่อนจากข้อมูลเรียนรู้ ข้อมูลกรณีพิเศษ (edge case) เป็นต้น รวมทั้งอาจพิจารณาให้มีกระบวนการ feedback loop[4] เพื่อนำผลลัพธ์มาปรับปรุงโมเดลให้มีความถูกต้องแม่นยำและน่าเชื่อถือต่อเนื่อง

(2.3) กรณีใช้ Generative AI ควรมีแนวทางลดความเสี่ยงจากการที่ระบบให้ข้อมูลที่เป็นเท็จ (hallucination) เช่น ใช้เทคนิค retrieval-augmented generation[5] เพื่อช่วยให้โมเดลมีแหล่งข้อมูลสำหรับอ้างอิง หรือปรับลักษณะคำตอบให้เหมาะสม เป็นต้น

---

[4] feedback loop คือ กระบวนการที่ผลลัพธ์หรือการตัดสินใจของโมเดลถูกนำกลับมาใช้เป็นข้อมูลเพื่อปรับปรุงและพัฒนาโมเดล

[5] retrieval-augmented generation (RAG) คือ เทคนิคปรับปรุงผลลัพธ์ของโมเดลประเภท Large Language Model ด้วยการให้โมเดลอ้างอิงฐานความรู้ที่เชื่อถือได้ซึ่งอยู่นอกเหนือจากแหล่งข้อมูลที่ใช้ในการเรียนรู้ก่อนที่จะแสดงผลลัพธ์ เพื่อควบคุมผลลัพธ์ของข้อความและสามารถเข้าใจวิธีการที่โมเดลสร้างคำตอบอย่างชัดเจนมากขึ้น (อ้างอิง: https://aws.amazon.com/what-is/retrieval-augmented-generation)

(2.4) มีแนวทางให้ผู้พัฒนา ผู้ใช้งาน และผู้ที่เกี่ยวข้องสามารถเข้าใจและอธิบายผลลัพธ์ของระบบ AI (explainability) ได้ เช่น การจัดทำเอกสารเพื่ออธิบายพารามิเตอร์ที่โมเดลใช้ การให้โมเดลประเภท Generative AI แสดงกระบวนการคิด (reasoning) หรือระบุแหล่งอ้างอิง (referencing) ควบคู่กับการแสดงผลลัพธ์ เป็นต้น

3) การควบคุมความเสี่ยงด้านไซเบอร์

ผู้ให้บริการทางการเงินควรมีแนวทางป้องกันและตรวจจับการโจมตีรูปแบบใหม่ที่เจาะจงกับระบบ AI รวมทั้งยกระดับแนวทางรักษาความมั่นคงปลอดภัยทางไซเบอร์ที่มีอยู่ให้ครอบคลุมการรับมือกับภัยดังกล่าว เช่น การทำ content filtering กับระบบ AI ทั้งขานำเข้าข้อมูล (prompt filtering) และขาแสดงผลลัพธ์ (response filtering) เพื่อป้องกันไม่ให้ระบบแสดงข้อมูลที่ไม่พึงเปิดเผยหรือข้อมูลเชิงลบ การทดสอบด้วยรูปแบบการโจมตีต่าง ๆ อย่างต่อเนื่อง เช่น ทดสอบการทำ data poisoning หรือ adversarial testing เพื่อหาข้อจำกัดและปิดช่องโหว่ของโมเดลที่ตรวจพบ

นอกจากนี้ ผู้ให้บริการทางการเงินควรติดตามและประเมินภัยคุกคามใหม่ ๆ อย่างต่อเนื่อง พร้อมปรับปรุงแนวทางป้องกันและตรวจจับให้ทันสมัยและเหมาะสมกับสถานการณ์ โดยอาจอ้างอิงจากมาตรฐาน เช่น OWASP Machine Learning Security Top 10 OWASP Top 10 for Large Language Model Applications เป็นต้น