生成式AI時代言論自由之憲法保障

——以AI生成內容管制為中心

國立中興大學法律系四年級

林芝筠

目次

壹、緒論……3

一、研究動機與目的……3

二、研究方法……3

貳、生成式 AI 與言論自由之基本概念……3

一、生成式 AI 之概念與發展……3

二、憲法言論自由之保障……4

三、生成式 AI 與言論自由之關聯……4

四、AI 生成內容之憲法定位……5

參、AI 生成內容對言論自由之挑戰……6

一、AI 生成內容與資訊真實性……6

二、深度偽造技術與言論自由……7

三、AI 生成內容與民主公共討論……8

四、AI 生成內容與其他基本權之衝突……9

肆、國家管制 AI 生成內容之憲法界限……11

一、國家管制之正當性……11

二、法律保留與法律明確性原則……12

三、比例原則之適用……13

四、事前審查與事後管制……15

五、不同類型 AI 生成內容之憲法審查……16

六、小結……18

伍、比較法與我國法制之建議……19

一、美國生成式AI與言論自由…… 19
二、歐盟生成式AI內容管制…… 20
三、比較法對我國之啟示…… 20
四、小結…… 21
陸、結論…… 21
一、研究發現…… 21
二、具體立法建議…… 22
參考資料…… 24

壹、緒論

一、研究動機與目的

近年生成式AI快速發展，ChatGPT等工具逐漸融入日常生活，使人民取得資訊、創作及表達意見的方式產生變化。生成式AI雖降低內容創作門檻，卻也衍生深度偽造影像、假訊息及輿論操縱等問題，對言論自由、民主公共討論及個人權益帶來新的挑戰。

我國憲法第11條保障人民之言論自由，因此，當生成式AI成為新興表意工具時，人民利用AI進行創作及表意是否受憲法保障，以及國家為防止AI遭濫用而進行內容管制時應受何種憲法限制，均值得探討。本文將以生成式AI生成內容為核心，分析其與言論自由之關聯，並從法律保留、法律明確性及比例原則等面向探討國家管制之界限，進一步參考國內外相關法制，提出兼顧科技發展、言論自由與公共利益之規範方向。

二、研究方法

本文採文獻分析法、法釋義學研究法、案例研究法及比較法研究法，蒐集憲法、行政法、資訊法及人工智慧法制相關文獻，分析我國憲法第11條言論自由之保障範圍、限制條件及違憲審查標準，並檢視AI生成內容相關法律規範及實務案例所涉及之基本權衝突，同時參考美國及歐盟相關法制，藉由比較不同規範模式，探討我國生成式AI內容管制之憲法界限及未來規範方向。

貳、生成式AI與言論自由之基本概念

一、生成式AI之概念與發展

人工智慧（Artificial Intelligence, AI）係指透過電腦系統模擬人類智

慧能力，以執行學習、推理、語言理解、辨識及決策等任務之技術。隨著機器學習及深度學習技術快速發展，人工智慧的應用已由傳統的資料分析與特定任務執行，逐漸擴展至能夠自主產製內容的生成式人工智慧（Generative AI）。[1]

生成式AI係人工智慧的一種類型，能依據使用者指令自動生成文字、圖片、聲音、影片及程式碼等內容，相較於傳統AI更具內容創作能力，已逐漸成為人民創作、取得資訊及表達意見的新興工具。

二、憲法言論自由之保障

我國憲法第11條規定：「人民有言論、講學、著作及出版之自由。」其中言論自由係我國憲法明文保障之基本權利，也是民主憲政秩序的重要基礎。且言論自由具有實現自我、溝通意見、追求真理及監督政治與社會活動等功能，並認為言論自由對於維持民主多元社會之正常發展具有重要意義，因此國家應給予高度保障。[2]

言論自由不僅保障人民表達意見的自由，也包含資訊取得、思想形成及參與公共討論等功能[3]。尤其在民主社會中，人民須能自由取得資訊並表達對公共事務的看法，方能監督政府及參與民主政治，因此言論自由具有高度的憲法保障價值。

三、生成式AI與言論自由之關聯

（一）生成式AI作為新興表意工具

傳統言論自由以自然人為表意主體，生成式AI則使人民得以透過AI快速創作文字、圖片、聲音及影片，成為新興表意工具。因此，應區分「AI本身是否享有言論自由」與「人民利用AI表意是否受言論自由保障」兩個問題。

---

[1] European Commission, AI Watch, Defining Artificial Intelligence 2.0, 2022

[2] 司法院大法官解釋第509號。

[3] 吳信華，憲法釋論，第五版，2024年9月，頁352-355。

有見解認為在特定脈絡下，AI 的演算輸出可以被理解為「機器表意」並受到表意自由保障[4]，惟本文認為，人工智慧本身並不當然具有基本權主體地位，但人民利用 AI 進行創作及表達思想之行為，原則上仍應受到憲法第 11 條言論自由之保障。

（二）AI 生成內容之表意主體問題

生成式 AI 不同於傳統被動傳遞資訊的工具，其可依使用者指令自主生成使用者事前未完全預期的內容，因而產生 AI 生成內容之表意主體究竟為何的問題。

本文認為，AI 應定位為人民表意的新興媒介，判斷其生成內容是否受言論自由保障，應依表意者、使用目的及內容性質具體判斷，不宜僅因內容由 AI 生成即排除於憲法保障之外。[5]

四、AI 生成內容之憲法定位

生成式 AI 所產生的內容種類十分廣泛，從單純娛樂性文字到政治評論、新聞資訊及藝術創作均可能涵蓋。因此，不宜將所有 AI 生成內容視為同一類型，並一概決定其是否受到憲法保障。

在憲法言論自由的架構下，應以內容本身之性質及其對基本權所具有的價值進行判斷。例如，人民利用 AI 創作小說、音樂或繪畫，主要涉及創作及表意自由；人民利用 AI 撰寫政治評論，則涉及民主公共討論；人民利用 AI 製作商業廣告，則涉及商業性言論[6]。且不同類型的言論具有不同程度的憲法價值，因此國家對其進行限制時，所應採取的審查標準亦可能有所不同。

---

[4] 陳柏良，AI 時代之分裂社會與民主　－以美國法之表意自由與觀念市場自由競爭理論為中心，月旦法學雜誌，第 302 期，2022 年，頁 118-121。

[5] 許炳華，新興溝通科技與言論自由——可能之定位暨思維，國立中正大學法學集刊，第 70 期，2021 年，頁 25-35。

[6] 參司法院大法官解釋第 414、577、744 號，認定商業性言論之性質。

參、AI 生成內容對言論自由之挑戰

一、AI 生成內容與資訊真實性

（一）生成式 AI 對資訊傳播之影響

傳統資訊傳播主要受到內容製作成本、時間及傳播媒介等因素限制，個人即使具有表意意願，亦可能受到製作技術及傳播能力的限制。然而，生成式 AI 降低了資訊產製的技術門檻，使個人得以透過簡單的文字指令快速生成大量內容，並透過社群媒體及網路平台迅速傳播。

此種技術發展對言論自由具有正反兩面的影響。一方面，生成式 AI 使更多人民能夠參與公共討論，降低資訊生產及表意的門檻，對言論市場的多元化具有正面意義；另一方面，當 AI 生成內容可以被大量且低成本地製造時，亦可能使錯誤資訊及虛假資訊以更快的速度進入公共討論。因此，生成式 AI 所造成的問題並非單純在於「假訊息」本身，而在於其改變了資訊產製與傳播的成本結構，使大量不實或難以查證的資訊可能在短時間內形成相當程度的社會影響。[7]

（二）AI 幻覺與資訊真實性問題

生成式 AI 另一項重要特徵，是其產生的內容並不必然與客觀事實相符。大型語言模型可能在缺乏可靠資料或無法確認答案時，產生看似合理但實際錯誤的資訊，此種現象通常被稱為「AI 幻覺」（AI hallucination）。此一問題在一般娛樂或創作情境下，未必直接涉及重大憲法問題，然而，若 AI 生成內容涉及公共政策、政治人物、選舉或其他公共事務，錯誤資訊可能進一步影響人民形成意見及參與民主政治的能力。

憲法法庭 112 年憲判字第 8 號判決即指出，在民主社會中，涉及公共利益的

[7] 許炳華，同註 5，頁 42-44。

事實性言論具有提供資訊及形成公共討論的重要功能，尤其透過電子網路傳播時，其影響範圍及反覆傳播的可能性均較大，因此表意人對於事實性言論應負有一定程度的查證義務[8]。

此一見解雖係針對誹謗罪所作成，但其對本文具有重要參考價值。因為生成式 AI 可能大量產生未經確認的事實性內容，若人民完全不經查證即將 AI 產出作為事實資訊散布，可能使資訊環境更加難以辨識真偽。然而，值得注意的是，要求人民查證 AI 生成內容，並不代表國家得以「資訊真實」為由全面限制言論。言論自由本即容許一定程度的錯誤與不確定性，因此不應僅因內容不實，即否定其憲法保障。

二、深度偽造技術與言論自由

（一）深度偽造技術之概念

深度偽造（Deepfake）係利用人工智慧生成或操弄影像、聲音及影片，使其高度模擬真實人物或事件，具有較高欺騙性。歐盟《人工智慧法》亦將深偽內容定義為足以使一般人誤認其真實性或來源之 AI 生成或操弄內容。[9]

（二）深度偽造技術對言論自由之正面功能

從言論自由的角度觀察，深度偽造技術並非當然應受到禁止。例如，人民可能利用 AI 製作政治諷刺影片、模擬歷史人物、進行藝術創作或以虛構方式評論公共議題[10]。此類內容雖然具有虛構性，卻仍可能具有思想表達、藝術創作及公共討論的價值。因此，若國家僅因內容係由 AI 生成，或具有虛構成分，即一律禁止，可能將合法的藝術創作及政治評論一併排除於言論自由保障之外。

---

[8] 憲法法庭：112 年憲判字第 8 號判決。

[9] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), Art. 3(60).

[10] 蔡佳蓁，AI 時代的資訊污染危機——刑法如何回應 Deepfake 技術下的數位剝削，2025 年，頁 19。

（三）深偽技術對言論自由之負面影響

深度偽造若被刻意用於欺騙，可能干擾人民形成政治意見及參與選舉[11]，並侵害他人之名譽、隱私及人格權[12]。因此，國家對於具有高度欺騙性且可能造成重大權益侵害的深度偽造內容，具有一定程度的管制正當性。然而，此種管制仍須符合比例原則，並應區分「具有虛構性但屬合法表意」與「以欺騙為目的且造成具體重大危害」的內容。

三、AI 生成內容與民主公共討論

（一）生成式 AI 與民主政治

言論自由不僅保障個人表達思想的自由，更具有維持民主公共討論及監督政府的重要功能。生成式 AI 一方面能協助人民理解公共議題、整理資訊，降低參與公共討論的門檻；另一方面，也可能被用於大量製造假訊息及政治深度偽造內容，影響人民對資訊真實性的判斷。由此可見，生成式 AI 所帶來的憲法問題，已不僅是「人民能否自由表達」，亦涉及「人民能否在適當的資訊環境中形成自由意見」。

（二）AI 生成內容與選舉資訊

選舉攸關民主政治，人民能否取得可信資訊，直接影響其政治判斷與投票決定。生成式 AI 所產生的深度偽造影片、偽造聲音及大量政治訊息，可能加劇選舉資訊操弄，尤其模擬候選人發言時，其欺騙性更為顯著。[13]

然而，若要求所有政治性 AI 內容均須經政府審查，恐使國家成為判斷政治言論真偽的權力者，並產生權力濫用及壓制批判言論之風險。因此，AI 政治內

---

[11] 林彥廷，深度偽造聲音影像妨害選舉公正性之立法評析，月旦律評，第 45 期，2025 年，頁 96。

[12] 同前註，頁 100-101。

[13] 蔡佳蓁，同註 10，頁 23-24。

容管制應避免以抽象的「資訊正確」為由，過度限制言論自由。

（三）資訊真實性與言論自由之平衡

憲法所保障的言論自由並不要求所有言論均為真實，民主社會的公共討論本即包含不同意見、推測、批判及錯誤判斷。憲法法庭 112 年憲判字第 8 號判決亦強調，涉及公共利益的事實性言論，其公益論辯貢獻度越高，原則上應給予越高程度的言論自由保障；同時，因網路及社群媒體具有高度傳播力，表意人亦應負擔一定程度的合理查證義務[14]。AI 生成內容不宜僅以真偽區分，應綜合考量言論性質、公共利益及危害程度。國家應防止的是具高度欺騙性及重大危害之資訊操弄，而非全面限制不實言論。

四、AI 生成內容與其他基本權之衝突

（一）言論自由與名譽權

生成式 AI 可能使人民更加容易製作涉及特定個人的虛假內容。例如利用 AI 生成某人從事違法行為的影片，或製作虛假的不當言論，均可能造成個人社會評價降低。此時即產生言論自由與名譽權之基本權衝突。

憲法法庭 112 年憲判字第 8 號判決認為，言論自由受憲法第 11 條保障，名譽權則屬憲法第 22 條保障之非明文權利；兩者均應受到憲法保障，當二者發生衝突時，應依個案情形進行適當利益衡量[15]。因此，若 AI 生成內容涉及他人名譽，不能僅因其屬於言論而當然受到絕對保障，也不能僅因其內容可能造成名譽損害，即完全否定言論自由。應依言論對公共討論的貢獻程度、內容真實性、傳播方式、損害程度及表意人主觀狀態等因素進行衡量。

尤其值得注意的是，AI 生成內容具有快速複製及大量傳播的特性，一旦涉

[14] 同註 8。
[15] 同註 8。

及名譽侵害，其損害可能迅速擴大。因此，傳統言論自由與名譽權的衡量標準，未來可能需要進一步考量 AI 生成內容所具有的規模性及傳播速度。

（二）言論自由與隱私權

生成式 AI 亦可能利用個人照片、聲音或其他資料產製高度擬真的內容。例如將個人的臉部影像套用至其他影片，或利用其聲音生成虛假的語音內容。此類情況可能同時涉及言論自由、隱私權、人格權及個人資訊自主等權利。

尤其當 AI 生成內容涉及私人生活領域，為私密敏感事項，且與公共利益無關時，國家對於保護個人隱私及人格尊嚴的正當性相對較高[16]；相反地，若內容涉及具有高度公共利益的政治人物或公共議題，則言論自由的保障價值可能相對提高。因此，AI 生成內容所涉及的隱私權衝突，也應依內容與公共利益的關聯程度進行個案判斷，而不能建立一個完全排除言論自由的絕對隱私權保障。

（三）言論自由與民主秩序之衝突

生成式 AI 最具特殊性的問題，在於其可能同時影響個人基本權及整體民主制度。例如，大量製作虛假的政治內容可能影響選民判斷，若國家完全不介入，可能使民主公共討論受到資訊操弄；但若國家以維護民主秩序為理由，廣泛限制政治言論，亦可能使政府取得過大的資訊控制權。[17]

因此，民主秩序本身既可能成為國家管制 AI 內容的正當目的，也可能成為國家限制政治言論時需要防範的權力濫用風險。在此情況下，國家管制的重點應從「判斷人民應該相信什麼」轉向「提高人民辨識資訊真偽的能力」。例如要求 AI 生成內容適當標示、提高來源透明度、建立內容溯源機制等，相較於直接刪

[16] 憲法法庭：111 年憲判字第 13 號判決。

[17] 陳柏良，同註 4，頁 112-113。1910 年代美國設置政治宣傳機構，並陸續訂定《間諜法》、《判亂法》對發言者及出版社課以刑罰，引發寒蟬效應。

除或禁止內容，對言論自由的侵害通常較低。

肆、國家管制AI生成內容之憲法界限

一、國家管制之正當性

（一）國家管制之必要性

生成式AI雖具有促進資訊流通及表意自由之功能，但其高度自動化及大量生成的特性，也可能造成假訊息、深度偽造內容、人格權侵害及民主程序遭受操弄等風險。因此，國家為保障人民權利及維護公共利益，並非完全不得對AI生成內容加以規範。

我國《人工智慧基本法》已於2026年1月14日公布施行，其立法目的包括建構人工智慧安全應用環境、保障人民基本權利、增進社會福祉及確保技術應用符合社會倫理。該法並要求政府推動人工智慧時，應遵循人類自主、人格權、隱私保護、透明及可解釋、公平及可問責等原則。[18]

由此可見，我國AI治理應兼顧科技創新與基本權保障。雖然國家對重大權利侵害具有管制正當性，但仍須具備法律依據，並符合法律明確性及比例原則。

（二）管制目的之憲法正當性

綜合以上零零總總生成式AI所生之影響及問題，國家對AI生成內容進行管制，可能基於下列目的：

第一，保護個人之名譽權、隱私權及人格權，例如防止利用深度偽造技術偽造他人影像或聲音。

第二，保障公共安全及交易秩序，例如防止AI生成內容被利用於詐欺或其

[18] 人工智慧基本法第1條、第4條。

他犯罪行為。

第三，維護民主公共討論及選舉制度，避免高度擬真的AI生成內容被用於大規模操弄選民。

第四，維護人民形成自主判斷所必要之資訊環境，避免資訊環境受到大量虛假內容干擾。

雖國家管制AI生成內容雖具有正當目的，但「維護資訊真實」不應成為全面限制言論的理由。民主社會本容許不同觀點及錯誤判斷，因此管制應以防止具體、重大之權利侵害為核心，而非消除所有錯誤資訊。

二、法律保留與法律明確性原則

（一）AI內容管制應有法律依據

國家若以行政處分、行政命令或其他公權力措施限制人民利用AI進行表意，涉及人民憲法上言論自由之限制時，自應具備適當之法律依據。尤其當管制措施涉及內容刪除、禁止發布、行政罰鍰或其他具有強制性的法律效果時，更不能僅透過行政機關發布之政策或內部規範加以限制。[19]

我國《人工智慧基本法》雖已建立AI治理的基本架構，但其性質主要在於確立人工智慧發展及治理的基本原則，並非所有AI生成內容均可直接依該法受到具體限制。換言之，未來若國家欲針對特定AI生成內容建立禁止或處罰規範，仍應視其對人民基本權之限制程度，由法律具體規範其構成要件及法律效果[20]。因此，AI治理基本法與個別作用法之間，應形成適當分工：前者確立治理原則，後者則針對具體風險建立明確之權利義務及責任規範。

---

[19] 司法院大法官解釋第443號。非屬對憲法第8條人身自由之限制之其他自由權係屬國會保留之事項，應以法律為之。

[20] 同前註。

（二）法律明確性原則

AI 生成內容管制另一項重要問題，是規範用語是否足夠明確。例如，若法律僅規定「不得生成不實資訊」、「不得生成有害內容」或「不得散布可能誤導社會之資訊」，人民可能難以預見何種內容會受到法律限制。

法律明確性原則的目的，在於使人民能夠預見自己的行為是否受到法律規範，並避免行政機關享有過大的解釋及裁量空間。司法院釋字第 636 號解釋即指出，若以法律限制人民權利，其構成要件應具體明確，使受規範者得以理解並預見其行為之法律效果，且應可經由司法審查加以確認，否則可能違反法律明確性原則。[21]

套用於 AI 生成內容，若法律以「假訊息」、「不當內容」或「有害資訊」等高度抽象概念作為處罰依據，卻未進一步說明判斷標準，可能使人民無法預測法律效果，並使行政機關取得過大的內容判斷權，可能產生裁量濫用之虞。因此，AI 生成內容之管制規範應盡可能具體說明其規範對象、構成要件、危害程度及法律效果，尤其涉及刑罰或重大行政處罰時，更應提高規範明確程度。[22]

三、比例原則之適用

（一）比例原則作為 AI 內容管制之核心審查標準

國家限制人民言論自由，應符合憲法第 23 條之比例原則。就 AI 生成內容而言，即使管制目的具有正當性，仍須進一步判斷所採取的手段是否適當、必要，以及對人民言論自由所造成的損害是否與其所追求的公共利益相當，即國家為達目的不得不擇手段，仍須符合一定比例。[23]

---

[21] 司法院大法官解釋第 636 號。
[22] 司法院大法官解釋第 690 號。涉及刑罰或重大行政處罰應從嚴審查法律明確性原則。
[23] 司法院大法官解釋第 791 號。

（二）適當性原則

首先，管制措施必須有助於達成其所追求的目的。比例原則之適當性要求，係指國家所採取之手段應具有促進或實現其所追求目的之可能性，若手段與目的之間欠缺一定關聯，即難以通過適當性之檢驗[24]。例如，國家要求深度偽造政治影片在發布時標示「AI 生成」或「經 AI 操弄」，其目的在於降低人民誤認內容為真實事件的可能性。此種透明化措施與降低資訊誤導風險之間具有一定關聯，因此原則上具有適當性。[25]

相反地，如果國家為防止假訊息，而直接禁止人民使用 AI 生成政治內容，卻無法證明此種全面禁止能有效降低資訊操弄風險，即可能難以通過適當性要求，亦即只有在對他人基本權利造成明顯及實質重大危害時，方具有正當性。[26]

（三）必要性原則

即使某項管制措施具有適當性，仍須進一步判斷是否存在侵害人民言論自由較小、但同樣有效的替代手段[27]。例如，國家若可以透過 AI 生成內容標示、來源揭露、平台通知、媒體識讀教育及事後救濟等方式降低深偽內容的危害，即不宜直接採取全面禁止或事前審查。[28]

此一問題尤其重要，因為 AI 生成內容具有高度多樣性，若採取一律刪除或禁止的方式，可能同時限制具有公共討論價值的政治評論、藝術創作及諷刺性內容。因此，對 AI 生成內容而言，原則上應優先採取對言論自由侵害較小的透明化及事後管制手段。

---

[24] 黃舒芃，比例原則及其階層化操作——一個著眼於司法院釋憲實務發展趨勢的反思，中研院法學期刊，第 19 期，2016 年，頁 14-17。
[25] 陳柏良，同註 4，頁 125-126。
[26] 林子儀，言論自由之理論基礎，台大法學論叢，第 18 卷第 1 期，1988 年，頁 273-275。
[27] 同前註。
[28] 許炳華，同註 5，頁 54-57。

（四）狹義比例原則

最後，即使管制措施具有適當性及必要性，仍須衡量其所造成的言論自由損害與所保護的公共利益是否相當[29]。例如，一段 AI 生成的政治諷刺影片雖可能具有誤導性，但如果一般人可以輕易辨識其為戲劇或諷刺作品，其對民主秩序的危害可能相對有限。若因此對創作者施以嚴重刑罰，可能造成過度限制。

反之，若有人於選舉前大量製作高度擬真的候選人深度偽造影片，並故意使選民相信候選人曾發表虛假言論，其對民主程序及選民自主決定的影響可能相當重大，此時國家採取較積極的管制措施，較具有正當性[30]。因此，AI 生成內容之管制不能僅依「是否為 AI 生成」作為判斷標準，而應考量內容性質、表意目的、傳播規模、欺騙程度及可能造成的實際危害。

四、事前審查與事後管制

（一）AI 生成內容之事前審查

事前審查係指言論在發布前即須經國家機關審查或取得許可。由於此種制度使國家得於人民表意之前決定言論能否進入公共領域，因此對言論自由具有高度干預性。司法院釋字第 744 號解釋即指出，言論之事前審查乃對言論自由的重大干預，原則上應認為違憲；除非立法資料足以支持該制度係為防止人民生命、身體或健康遭受直接、立即且難以回復之危害等特別重要公共利益，並具有直接且絕對必要之關聯，始可能符合比例原則[31]。學者指出本號解釋對言論自由事前審查採取較為嚴格之審查標準，並要求相關管制標準具有具體、明確性，同時應提供人民即時之司法救濟機會。[32]

---

[29] 司法院大法官解釋第 699 號林錫堯大法官協同意見書。
[30] 林彥廷，同註 11，頁 99-103。
[31] 司法院大法官解釋第 744 號。
[32] 楊智傑，言論自由事前審查之審查標準——釋字第 744 號解釋與美國審查標準比較，憲政時

因此，若國家要求所有AI生成內容在發布前均須經政府審查，將面臨高度的憲法疑慮。尤其政治性AI生成內容涉及公共討論時，若由行政機關決定何種內容可以發布，可能使國家成為公共言論的「守門人」，進而產生政府壓制批判性言論的風險。

（二）事後管制之優先性

相較於事前審查，事後管制係允許人民先行表意，待言論造成法律所禁止之具體侵害時，再透過民事、行政或刑事責任進行處理。此種方式原則上較能兼顧言論自由與公共利益。

例如，對於AI生成的名譽侵害內容，可透過民事損害賠償、除去侵害及其他法律救濟處理；對於涉及詐欺或其他犯罪的AI生成內容，則可依既有刑事法律追究責任。因此，除非AI生成內容所造成的危害具有高度即時性、重大性且難以透過事後救濟回復，否則國家原則上應優先採取事後管制，而非全面性的事前審查。此種以較低侵害之手段處理AI生成內容之方式，也與必要性原則要求國家優先採取對基本權侵害較小手段之意旨相符。

事後管制優先並非表示絕對禁止事前管制。若AI生成內容可能造成高度即時、重大且難以回復之危害，例如選舉期間散布高度擬真之政治深度偽造，國家仍可能採取較積極措施[33]。惟原則上應優先採取內容標示、來源揭露及事後救濟等較低侵害手段之措施[34]；僅於具體危害具有高度即時性、重大性及難以透過事後救濟回復之特性，且有充分證據證明事前管制確有必要時，始可能例外採取較為積極之事前限制。

代，第43卷第1期，頁99-100。

[33] 林彥廷，同註30。

[34] 林彥廷，同註30。

五、不同類型 AI 生成內容之憲法審查

生成式 AI 所產生的內容具有高度多樣性，因此不宜採取單一的管制標準。本文認為，應依言論內容及可能產生的危害程度，建立分級的憲法審查模式。又如前所述，應以事後審查為優先，故以下之審查密度係針對不同言論類型所為之事後審查。其中對言論內容所為之限制，原則上應適用最嚴格之審查方式，尤其應避免以「內容不正確」作為廣泛限制的理由，合先敘明。[35]

（一）政治性及公共議題之 AI 生成內容

涉及政治意見、公共政策及政府監督的 AI 生成內容，與民主公共討論具有高度關聯，因此應受到較高程度的言論自由保障[36]。故於審查標準之選擇上，應適用最為嚴格之審查密度，即係為追求憲法上重大公益，且手段與目的間須具絕對必要關聯。

（二）藝術、創作及娛樂性內容[37]

AI 生成小說、圖片、音樂、電影或諷刺作品，主要涉及創作及表意自由。即使內容為虛構，也不代表其缺乏憲法價值。因此，國家對此類內容原則上應採取較低度的干預方式，例如要求適當揭露 AI 生成性質，而非直接禁止內容，故於審查限制合憲性時亦須採取嚴格之審查標準。

（三）商業性 AI 生成內容

AI 生成廣告及商業宣傳雖受到言論自由保障，但其主要目的在於促進交易，因此與政治性言論相比，其憲法保障強度可能有所不同[38]。相關學理亦指出，我國商業言論之憲法保障及其限制，應依其所涉及之利益及管制目的進行具體判斷

---

[35] 司法院大法官解釋第 445 號。針對言論內容之限制，於審查合憲性時，應提高審查密度。
[36] 司法院大法官解釋第 414 號。政治性言論屬高價值言論。
[37] 同註 35。藝術性言論屬高價值言論。
[38] 同註 36。商業性言論非關公意形成，屬低價值言論，國家對之可採取較為高度之限制措施

[39]。若 AI 生成廣告具有虛偽、誤導或欺騙消費者之情形，國家得基於消費者保護及市場交易秩序進行較積極的管制。因此在限制合憲性之審查依個案情況可能採取中度審查，亦或係嚴格審查。

（四）涉及人格權侵害之 AI 生成內容

若 AI 生成內容涉及偽造他人影像、聲音或私人生活內容，並造成名譽、隱私或人格權侵害，國家具有較強的管制正當性。此時，應透過言論自由與其他基本權的利益衡量，判斷是否有必要採取內容移除、損害賠償或其他法律責任。

憲法法庭 112 年憲判字第 8 號判決亦指出，言論自由與名譽權均具有憲法保障地位，發生衝突時應依言論對公益論辯的貢獻程度，以及名譽權受侵害的方式、程度及範圍進行衡量。[40]

（五）附論

前（一）至（三）所述之言論類型之審查皆係針對言論內容之事後限制，若係非針對言論內容之事後限制，原則上於合憲性審查時採中度審查標準即可，惟若該限制難以使用雙軌區分，且該限制係為使表意人無法發表特定內容或使公眾曉諭，實質上與對言論內容的限制無異，仍須提升審查密度[41]；又對於事前審查之部分，已於前第四小節敘明對言論內容之事前審查原則違憲，僅在例外情況下方有合憲可能，且須適用最嚴格之審查標準，反之，非針對言論內容之事前審查採中度標準即可。[42]

六、小結

綜上，國家基於保障人格權、公共安全及民主秩序等重要公益，對 AI 生成

---

[39] 賴祥蔚，商業言論與憲法的言論自由保障，臺灣政治學刊，第 14 卷第 1 期，2010 年，頁 185-190。
[40] 同註 8。
[41] 憲法法庭 112 年憲判字第 17 號判決黃昭元大法官協同意見書第 11-12 段。
[42] 司法院大法官解釋第 806 號。

內容進行管制具有一定正當性，但若涉及人民思想表達及資訊傳播，仍應受到言論自由之保障。因此，相關規範除須具備法律依據並符合法律明確性原則外，亦應通過比例原則之檢驗。

此外，考量事前審查對言論自由具有高度干預性，國家原則上應優先採取內容標示、來源揭露及事後責任等侵害較小之手段，並依AI生成內容之性質、言論價值及可能造成之危害程度採取差異化管制。如此方能在防止AI遭濫用與保障人民言論自由之間取得適當平衡。

伍、比較法與我國法制之建議

一、美國生成式AI與言論自由

美國憲法第一修正案高度保障言論自由，尤其重視政治性言論及公共議題之討論。美國聯邦最高法院於 *United States v. Alvarez* 一案指出，虛假陳述並非當然不受第一修正案保障，國家若欲加以限制，仍須具備正當理由及必要性。[43]

此一見解對生成式AI具有重要啟示。AI生成內容即使具有不實性，亦不能僅因「不真實」即排除於言論自由保障之外。若國家以資訊真實性作為全面管制依據，尤其涉及政治人物、政府政策或選舉時，可能使政府取得過大的言論內容判斷權。

此外，AI深偽技術可能製作高度擬真的政治人物影音，影響選民判斷。因此，美國近年逐漸思考如何在防止選舉資訊遭操弄與保障政治言論之間取得平衡，而非全面禁止政治性AI生成內容。[44]

---

[43] United States v. Alvarez, 567 U.S. 709 (2012).

[44] Sophie Loewenstein,Make America F e America Fake Again?: Banning Deepfak e Again?: Banning

二、歐盟生成式AI內容管制

相較於美國，歐盟對人工智慧採取較積極的風險治理模式，強調基本權利、透明度及社會安全。歐盟於2024年通過《人工智慧法》(Artificial Intelligence Act, AI Act），建立風險分級的治理架構，並針對生成式AI及深度偽造內容設有透明度要求。[45]

AI Act第50條規定，特定AI生成或操弄的文字、圖片、聲音及影片應具備可辨識的標示；對於深度偽造內容，亦要求揭露其人工生成或操弄性質[46]。由此可見，歐盟並非全面禁止AI生成內容，而是透過透明標示降低資訊誤導風險。

此外，AI Act亦考量藝術、創作、諷刺及虛構作品，要求透明義務之適用避免不當妨礙作品呈現及利用[47]。因此，歐盟模式對我國的啟示在於，面對AI生成內容所造成的風險，可優先採取透明標示等較低侵害性的手段，以兼顧資訊安全與言論、創作自由。

三、比較法對我國之啟示

綜合美國與歐盟的制度，可以發現兩者雖採取不同的管制模式，但均未將AI生成內容一律排除於言論自由保障之外。美國較重視第一修正案對政治性言論的高度保障，避免國家以「資訊不實」為由過度限制言論；歐盟則透過風險分級及透明標示降低AI生成內容可能造成的危害。

對我國而言，生成式AI內容管制不宜採取全面禁止或一律事前審查，而應依內容性質及可能造成的危害程度採取分級管制。對於政治性及公共議題之內

Deepfakes of Federal Candidates in Political Advertisements Under the First Amendment,93 Fordham L. Rev.2024,306-309.

[45] 同註9。

[46] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), Art. 50.

[47] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), Art. 50(4) and Recital 134.

容，應給予較高程度的言論自由保障；對於深偽、詐欺或侵害人格權等具有較高危害性的內容，則可採取較積極的管制措施。

此外，歐盟所採取的AI生成內容標示制度亦值得我國參考。相較於直接禁止內容，透過揭露AI生成性質，使人民自行判斷資訊真實性，對言論自由的限制較小，也較符合比例原則。因此，我國未來AI治理應以風險分級、透明標示及事後救濟為主要方向，在防止AI遭濫用的同時，避免國家過度介入人民的言論及創作自由。

四、小結

綜合前述比較可知，美國與歐盟雖採取不同的AI治理模式，但均重視在防止AI濫用與保障言論自由間取得平衡。美國強調避免政府過度限制政治性言論，歐盟則透過風險分級及透明標示降低資訊誤導風險。對我國而言，未來生成式AI管制應避免全面禁止或事前審查，並依內容性質及危害程度採取適當措施，以兼顧言論自由、資訊安全與公共利益。

陸、結論

一、研究發現

本文以生成式AI與言論自由之憲法保障為核心，分析人民利用生成式AI進行創作、資訊取得及意見表達所涉及之基本權問題。研究發現，生成式AI雖具有高度自主生成內容之能力，但人工智慧本身並非當然具有基本權主體地位，人民利用AI進行表意之行為，原則上仍應受到憲法第11條言論自由之保障。因此，國家不應僅以「AI生成」作為限制或禁止內容之理由。

其次，生成式AI所造成的假訊息、深度偽造及人格權侵害等問題，確實使

國家具有一定的管制正當性。然而，言論自由並不要求所有言論均為真實，尤其政治性及公共議題之言論具有高度民主價值，國家更應避免以「維護資訊真實」為由進行過度管制。對於AI生成內容之限制，仍須符合法律保留、法律明確性及比例原則，並依內容性質、表意目的、傳播方式及可能造成的危害程度進行個別判斷。

故本文認為，AI內容管制應以「降低危害」而非「消除所有錯誤資訊」為主要目標。比較美國及歐盟法制可知，美國強調政治性言論之高度保障，歐盟則透過風險分級及透明標示降低資訊誤導風險。此均顯示AI治理並非只能在「放任」與「全面禁止」之間二擇一，而應尋求兼顧言論自由與公共利益之中間途徑。

二、具體立法建議

第一，建立AI生成內容之風險分級制度。立法上不宜將所有AI生成內容一律納入相同管制，而應依內容性質及危害程度區分。對於政治性、公共議題及藝術創作等具有高度言論價值之內容，原則上應採取較低度的限制；對於涉及詐欺、重大人格權侵害或高度欺騙性深度偽造內容，則得採取較積極之管制措施。

第二，建立AI生成內容之透明標示制度。對於具有高度擬真性且可能造成誤認之AI生成或操弄內容，可要求適當揭露其AI生成性質，使資訊接收者能夠自行判斷，而非由國家直接決定內容是否得以發布。相較於全面禁止，此種方式對言論自由之侵害較小，亦較符合比例原則。

第三，限制政府對政治性AI內容之事前審查。涉及政治人物、選舉及公共政策之AI生成內容，應受到較高程度之言論自由保障。除非內容具有高度即時、重大且難以回復之危害，否則不宜建立一般性的政府事前審查制度，以避免行政機關成為政治言論的「守門人」。

第四，強化AI生成內容之事後救濟機制。對於AI生成內容所造成之名譽、隱私及人格權侵害，可透過損害賠償、內容移除及其他法律責任予以處理；對於涉及詐欺或其他犯罪之情形，則應由既有刑事法律加以規範。透過事後責任取代全面事前管制，可降低對合法言論及創作活動之影響。

第五，明確規範AI內容管制之法律要件。未來若制定專門規範，應避免使用「不當內容」、「有害資訊」等過於抽象之概念，而應具體規定受規範之內容類型、危害程度、表意目的及法律效果，以避免人民無法預見法律效果，並防止行政機關享有過大的裁量空間。

總結而言，我國未來生成式AI立法應以「風險分級、透明標示、事後救濟及限制事前審查」為主要方向，在保障人民言論自由的前提下，針對具有具體且重大危害之AI生成內容進行必要管制。如此不僅能降低生成式AI遭濫用所產生的社會風險，也能避免國家因追求資訊真實而過度介入人民表意，進而在科技創新、基本權保障與民主公共利益之間建立較為適當的平衡。

參考文獻

一、司法院大法官解釋、憲法法庭

司法院大法官解釋第 414、445、509、577、636、690、744、791 號。

司法院大法官解釋第 699 號林錫堯大法官協同意見書。

憲法法庭 111 年憲判字第 13 號判決。

憲法法庭 112 年憲判字第 8 號判決。

憲法法庭 112 年憲判字第 17 號判決黃昭元大法官協同意見書。

二、專書

吳信華，憲法釋論，第五版，2024 年 9 月。

三、期刊

林子儀，言論自由之理論基礎，台大法學論叢，第 18 卷第 1 期，1988 年。

林彥廷，深度偽造聲音影像妨害選舉公正性之立法評析，月旦律評，第 45 期，2025 年。

許炳華，新興溝通科技與言論自由——可能之定位暨思維，國立中正大學法學集刊，第 70 期，2021 年。

陳柏良，AI 時代之分裂社會與民主－以美國法之表意自由與觀念市場自由競爭理論為中心，月旦法學雜誌，第 302 期，2022 年。

黃舒芃，比例原則及其階層化操作——一個著眼於司法院釋憲實務發展趨勢的反思，中研院法學期刊，第 19 期，2016 年。

楊智傑，言論自由事前審查之審查標準——釋字第 744 號解釋與美國審查標準比較，憲政時代，第 43 卷第 1 期。

賴祥蔚，商業言論與憲法的言論自由保障，臺灣政治學刊，第 14 卷第 1 期，2010

年。

四、學位論文

蔡佳蓁，AI 時代的資訊污染危機——刑法如何回應 Deepfake 技術下的數位剝削，2025 年。

五、外國文獻

（一）美國

United States v. Alvarez, 567 U.S. 709 (2012).

Sophie Loewenstein, Make America Fake Again?: Banning Deepfakes of Federal Candidates in Political Advertisements Under the First Amendment, 93 Fordham L. Rev. (2024).

（二）歐盟

European Commission, AI Watch, Defining Artificial Intelligence 2.0 (2022).

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), Art. 3(60).

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), Art. 50.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), Art. 50(4) and Recital 134.