

法律人工智能的論證功能與模型建構

封明坤*

摘要 一般來說，論證的功能有三：證成、反駁與說服。傳統基於形式邏輯建模的智能論證系統忽略了法律論證的似真性、辯證性與語用性，因而僅能實現論證的證成功能。非形式邏輯為法律智能論證系統的建模提供了新的理論來源，彌補了形式論證功能的不足。本文試圖將非形式邏輯應用於法律人工智能的論證模型，通過“可廢止性”啟發法律人工智能從對話性和論辯性的視角來詮釋法律論證的反駁功能；通過非形式邏輯的語用維度，構建起智能論證系統的“可接受性”框架，將說服目標聽眾作為評價論證的充分性標準。除此之外，證據推理與價值衡量中的主觀因素作為模型建構中的重要考量，提醒我們應意識到智能論證模型的運行限度，並在特定環節進行適當的人為干預。

關鍵詞 法律人工智能 論證功能 非形式邏輯 論證模型

一、引言

1956年，“人工智能”（Artifical Intelligence）一詞在達特茅斯會議上被首次提出，自此人工智能成為一門獨立的學科逐漸走進大眾的視野，並逐漸廣泛應用於各領域。在法學領域，為減輕法官審判負擔並限制恣意裁判，法律人工智能的研究愈發深入。目前來看，法律人工智能的研究方向可以分為兩種。一種可以稱之為法與人工智能，致力探究面向人工智能新發現的法律應對。另一種則基於二者在應用上的融合，從模型建構的角度討論適用於法律領域的人工智能系統。進一步地，法律人工智能系統的開發從技術上又可區分為聯結主義路徑與符號主義路徑。聯結主義試圖賦予計算機以人類的心智觀，通過人工神經元網絡系統及深度學習等操作實現人工智能的自主決策，進而使人工智能能夠“像法官一樣思考”；相反，符號主義並不試圖去發掘人類心靈的底層邏輯，而是試圖在計算機上實現邏輯的推演。這一路徑將法律論證的邏輯建模視為法律人工智能系統構建的核心問題，並發展出了諸多優秀的司法專家系統。

本文將沿著符號主義的研究路徑討論法律人工智能的論證模型。堅持這一路徑的原因並非否定聯結主義在法律人工智能研究中的突出成果，而是符號主義更能充分還原現實的司法運作。從法學的角度來看，司法審判並不會僅僅關注結論的產生，更多的是法官對結論提供論據並進行論證說理

* 封明坤，中山大學法學院博士研究生。

的過程。機器學習的方法所採用的是關聯驅動從大量數據中擬合並總結出相關規律，但是由於缺乏人類的推理能力，其很難給出數據中的因果關聯。^[1]符號主義的進路可以較為清晰地還原法律推理與論證的具體過程，只不過需要採用一種從形式邏輯到非形式邏輯的思維轉換。本文認為，這一邏輯基礎的轉換能夠提供一種法律人工智能論證模型的構建思路，更重要的是幫助我們重新審視法律論證的邏輯刻畫方式，並以更加精細化的方式發掘法律論證的證成機理。

下文將首先從理論上闡明這一邏輯轉換的必要，其次基於非形式邏輯的可廢止性與語用特徵嘗試構建起一種智能論證模型，它包含了法律論證的論辯結構及其“可接受性”框架。最後將對模型建構的其他考量因素進行分析。

二、法律人工智能論證的邏輯基礎：從形式邏輯到非形式邏輯

基於符號主義的構建路徑關注到了智能論證模型與法律論證實踐相擬合的重要性，但傳統思路大多建立在形式邏輯的基礎上，這使其難以完全呈現法律論證的自然屬性與特徵。究其原因，可以認為形式邏輯無法滿足法律論證的似真性、辯證性、語用性特徵，而這又可從根本上歸結於法律的不完備性、不一致性。我們在此試圖論證的是，非形式邏輯源於法律論證的實踐情境，在研究對象、關注的論證要素、評價維度方面不同於形式邏輯的運用，因而能夠更為全面地還原法律論證的本身面貌，並為人工智能論證模型的構建提供方法支撐。

（一）形式邏輯的功能缺位與成因

關於“論證”，語用論辯理論的提出者范愛默倫（Frans H.van Eemeren）、荷羅頓道斯特（Rob Grootendorst）等人曾給出過一個準確的定義：“論證是一項關涉社會性、言辭性、理性的活動，旨在通過陳述來證成自己立場的可接受性，從而說服理性的批判者”。^[2]這一定義總結出了論證的證成性、反駁性、說服性功能，並將論證從兩個層面予以區分。第一個層面是“作為結果的論證”（argument-as-product）。此時，論證僅作為一個包含結論命題與前提命題的命題序列，強調論證的證成性。第二個層面是“作為過程的論證”（argument-as-process）。它更多地將論證視為論證者為使自己的主張被聽眾所接受而進行互相論辯的過程，因而更具動態性、目的性、多主體性、語境敏感性、對話性等特點。^[3]

形式論證通過形式邏輯對自身內部有效性標準進行考察。演繹論證是較為典型的形式論證方式，在前提為真的情況下通過演繹論證必然推出結論也為真。亞里士多德給出的經典例子是：凡人皆有一死；蘇格拉底是人；因此，蘇格拉底也會死。傳統法律人工智能論證系統大多基於形式邏輯進行設計，基於規則與案例的“專家系統”是這一路徑的主要成果。^[4]此路徑作為早期人們探索法律人工智能系統的主要方式雖然取得了很大發展，但囿於形式邏輯的適用局限，其仍舊無法處理諸

[1] 參見丁夢遠、蘭旭光、彭如、鄭南寧：《機器推理的進展與展望》，載《模式識別與人工智能》2021年第1期，第2頁。

[2] Frans van Eemeren, Rob Grootendorst & Francisca Snoeck Henkemans, *Argumentation: Analysis, Evaluation, Presentation*, Lawrence Erlbaum Associates, 2002, p. xii.

[3] 參見熊明輝：《訴訟論證——訴訟博弈的邏輯分析》，中國政法大學出版社2012年版，第24-27頁。

[4] 例如，1981年，沃特曼（Donald Waterman）和彼得森（Mark Peterson）基於規則推理的建模建立起了用以解決產品侵權責任法律裁定的專家系統；阿什利（Kevin D.Ashley）與其導師里士蘭（Edwina L.Risslan）則基於案例推理開發了海波（HYPO）系統。see Rissland Edwina L. et al, *AI and Law: A Fruitful Synergy, Artificial Intelligence*, 2003, pp.1-15. see Kevin D.Ashley, *Reasoning with cases and hypotheticals in HYPO*, International Journal of Man-Machine Studies, 1991, p.753-796.

多法律實踐場景。根本原因在於，基於形式邏輯所刻畫的論證模式僅僅滿足我們對論證第一層面的要求，其將法律視作一個封閉且靜態的規則體系，人工智能僅需要把特定的案件要素數據對應於特定的法律後果，進而將法律意義上的“證成關係”轉化為算法操作的“對應關係”。正如韋伯（Max Weber）所最早描述的那樣，他將法官通過形式理性裁判得出結論的過程形象地比作“自動售貨機”，認為法律體系應當是確定而又非人格化的。其尤為關注形式理性的法律，因為這代表著抽象而又高度邏輯化、“制度化”的普遍思維，而“制度化代表了一種邏輯清晰、一以貫之的完美規則體系（至少在理論上）”。^[5]

返回到法律本身，這種規則演繹的封閉算法背後實際上蘊涵的是機械主義的法律觀。拋卻其他爭議，韋伯的法律觀點的確為法律人工智能系統的構建提供了一種理論上的支撐，並且使得首架基於規則推理模式的法律人工智能從理論變為現實，但其終究是無法用來恰當描述法律論證的性質。法律自身的“不完備性”（incompleteness）與“不一致性”（inconsistency）是導致這一事實的根本原因，具體來說：

法律的不完備性體現在語言的模糊性、規則的例外情況及法律自身的漏洞等等。哈特將導致這些實踐困境的原因歸結為“法律規則的開放性結構”，認為法律規範自身所具有開放結構本質上導致了適用中的不確定性。在他所描述的“禮堂脫帽”的例子中，“任何人進入教堂時都應當把帽子摘下”作為一種規則的指引方式，一般來說，“人們只需確定某個語詞所能夠囊括的特殊事例，並將待定事實‘涵攝’在一般分類之下，便能夠通過簡單的三段論結構得出結論。”^[6]然而，究竟是用左手摘帽還是右手摘帽？帽子摘下後放在什麼地方？這些都是不清楚的。可見，現實情況中特定的事實案例並不會打上標籤以證明自己具有某些語詞所指代的特徵，規則本身也不會主動“承認”自己所包含何種事例。因此，邊緣事實的不確定性蔓延至規則中時，不管是立法還是判例，不論它們如何順利地適用於日常案例，當其解決邊緣案件時將會變得尤為難以把握。進而，通過形式邏輯所力求實現的嚴格“對應關係”也會淪為不切實際的幻想。

除此之外，法律的不一致性也會導致形式邏輯脫離論證實踐的要求，這種不一致性體現在法律規範間的衝突。毫無疑問，法律規範本身應當是內部和諧且協調的，意味著在進行法律論證的時候相互衝突的前提命題不能同時為真。但事實上，同一法律內部及不同法律之間的規範衝突時常會發生。

法律自身的“不完備性”及“不一致性”使得法律的論證活動具備特殊的性質，簡單的形式邏輯無法充分滿足其特性。這種特質表現在三個方面：似真性、辯證性、語用性。

首先，“似真性”（plausibility）有時也被稱為“合情性”。^[7]傳統邏輯通常認為，論證結論是否為真與論證形式密切相關——演繹方法中的前提為真，則結論必然為真。而似真性的思想則為規則的例外留出空間，將支持力放在命題的背後，只給出暫時接受某一命題的基礎。沃爾頓（Douglas N. Walton）將似真推理看作是區別於演繹與歸納的第三種類型，這一推論的特徵是：如果前提為真，那麼結論似然為真，^[8]其建立在人類經驗常識和自然感知的基礎之上。^[9]作為法律判斷中小前

[5] 參見[德]馬克斯·韋伯：《論經濟與社會中的法律》，張乃根譯，中國大百科全書出版社1998年版，導論第22-30、61頁。

[6] [英]哈特：《法律的概念》，張文顯等譯，中國大百科全書出版社1996年版，第125頁。

[7] 參見熊明輝：《訴訟論證——訴訟博弈的邏輯分析》，中國政法大學出版社2012年版，第60頁。

[8] 參見[加]道格拉斯·沃爾頓：《法律論證與證據》，梁慶寅、熊明輝譯，中國政法大學出版社2010年版，第111頁。

[9] 熊曉彪：《論似真推理的法律適用》，載陳金釗、謝暉主編：《法律方法》2024年第3期，法律出版社2024年版，第27頁。

提的案件事實便具有“似真性”，從證據到事實的推論過程本身具有科學性與猜測性相統一的特徵。^[10]而形式邏輯對“似真性”的忽視導致其缺乏塑造論證反駁功能之基礎。

其次，法律論證是“辯證的”（dialectical）。亞里士多德在《論題篇》中描述了論辯中的辯證特徵：論辯應當是一個具有目的性的過程，這一過程至少要包含兩名參與者，並且圍繞命題的提出與反駁而展開。^[11]法律論證的進程正是在原被告之間的不斷質疑中向前推進的，同時，法律論證的參與者當然也秉持各自的目的，即通過爭辯而駁倒對方的訴訟請求轉而正當化自己的訴訟請求。形式邏輯囿於其單主體的論證結構，在缺乏“似真性”考量的基礎上，更難以還原法律論證的反駁功能。

傳統形式邏輯不能充分描述法律論證的最後一個原因在於其忽視了論證是“語用性的”（pragmatic），這使其無法生動還原法律論證的說服功能。法律論證中的特定句式和表述通常不是單純語言的編碼意義，並不能從字面義直接得到聽眾想要獲得的意義，^[12]因此辨者對聽眾的說服都需要借助言說的意義推動才能得以實現。例如，法律條文中的“故意”與“過失”顯然不能僅根據語言習慣進行理解。同時，論證的語用評價標準具有語境依賴性，這意味著法律論證絕不是一個“非真即假”的判定過程。由於不同情境下所包含的材料不同，某一情境下的事實材料可能支持結論，而另一情境也可能會阻止同一結論的得出。

（二）非形式邏輯對論證功能的補充

由此觀之，法律論證是關涉對話、意見分歧、目標、聽眾、語境諸因素的複合體。^[13]目前而言，基於傳統形式邏輯建模尚不能滿足法律論證的所有特性，因而無法實現智能論證模型的反駁與說服功能。與之相較，非形式邏輯實際並非站在形式邏輯的絕對對立面，而是對形式邏輯進行論證評價時的必要補充，^[14]其在研究對象、所關注的論證要素、評價標準方面皆與形式邏輯不同。這導致非形式邏輯的優勢在於，能夠通過邏輯維度的可廢止性回應法律論證的似真性特徵，並在此基礎上將論證主體、論證目的等對話性要素考慮在內，還原庭審主體間（起方、應方、審方）的論辯狀態。同時，它將語用維度作為論證評價的標準，兼顧法律論證的情境依賴性，進而彌補形式邏輯在智能論證模型建構方面的功能缺位。具體而言：

第一，非形式邏輯通過可廢止性回應法律論證的似真性。傳統形式邏輯的研究對象是“蘊涵”，描述的是命題之間的邏輯關係。典型的範例就是演繹的推理形式，前提為真是演繹推理的一個要件。但正如先前所述，法律論證的大小前提本身就具有不確定性，因此難以刻畫法律推理的全過程。“可廢止”是指“概念的初顯性適用在例外情況下會被終止”。^[15]這意味著非形式邏輯並不需要從一個確定無疑的前提展開，當原本妥當的法律論證因為新出現的理由而變得不妥當的時候，可廢止性允許對推理的前提進行必要修正。這在系統的建模問題上可以對規則的“例外”做出恰當的處理。例如法律規則中的“但書”結構，當例外情況得到滿足時，原本法律的適用就會遭到阻卻。

[10] 參見杜文靜：《論刑事證據推理的似真性》，載陳金釗、謝暉主編：《法律方法》2015年第2期，山東人民出版社2015年版，第184頁。

[11] See J. Anthony Blair, Ralph H. Johnson, *Argumentation as dialectical, Argumentation*, 1987, p.45-47.

[12] 參見[法]丹·斯珀波、[英]迪埃珏·威爾遜：《關聯：交際與認知》，蔣嚴譯，中國社會科學出版社2008年版，第3頁。

[13] 參見陳鑫泉、武宏志：《非形式邏輯的理論成就及其對法律論證的意義》，載《湖南科技大學學報（社會科學版）》2014年第4期，第39頁。

[14] 參見熊明輝：《訴訟論證——訴訟博弈的邏輯分析》，中國政法大學出版社2012年版，第12頁。

[15] 參見董焱焱：《圖爾敏模型在法律論證中的應用與深化——以于歡案為應用實例》，載陳金釗、謝暉主編：《法律方法》2022年第1期，中國出版集團2022年版，第174頁。

第二，非形式邏輯可以在可廢止性的基礎上實現論證的反駁功能。顯而易見，“證成”與“反駁”是截然不同的：前者是論證的“正向建構”，一旦結論被推出，就不會被輕易動搖；後者則是一種對抗或限制的理由，它引入了“結論的可撤銷性”，即推理並非一旦成立就永遠有效。由於傳統的形式邏輯通常是單調邏輯（monotonic logic），這意味著論據的增加僅會增加結論，而不會推翻已有結論，因此在人工智能的建模過程中很難實現對原有結論的廢止。但是，現實的法庭論辯一定是建立在“證明——反駁”的循環往復之上的，庭審的主體從不僅限法官一人，判決的產生也絕不是法官的獨斷。缺乏攻擊與反駁的論證一定是“零主體”的“無目的的”。正如希契柯克（David Hitchcock）所總結的那樣，他認為論證是為論證者根據個人目的，通過提出支持主張的理由來說服聽眾的過程，這一過程應當包含相互反駁的辯證層。^[16]沃爾頓亦認為，論辯的過程應當允許質疑，即允許對方提出相反意見；現實法庭辯論亦復如此，法官不可能僅聽取一方之言就得出判斷。^[17]而非形式邏輯通常是非單調的（non-monotonic），這種非單調性意味著除了證成之外，原有的結論也可能因為新的信息而被廢止。非形式邏輯正是將這種可廢止性引入更廣闊的對話場景中進行理解，使其能關注論證要素中的“論證主體”與“論證目的”。因此，可廢止性體現在論證主體之間有目的的質疑與辯護的動態變化上，進而啟發法律人工智能從對話性與論辯性的視角來詮釋法律論證的反駁功能。

第三，非形式邏輯引入語用的評價維度，進而實現論證的說服功能。鑑於邏輯與語言的內在聯繫，對語言研究維度的不同側重導致了論證評價標準的不同傾向。人工語言是形式邏輯研究的語言載體，是指為達成某一目的而人工自行設計的語言；在這一研究旨趣上，形式邏輯僅關注語言的語形、語義維度。前者描述符號與符號之間的關係，後者則重視符號及其所指之間的關係。非形式邏輯對此的理解並不相同，它更關注經由人類自身發展所形成的自然語言，認為論證應當是對特定人展開的目的性說服，因此，論證需從語用的維度出發，重點處理符號使用者與符號之間的關係。二者之間的區別實際上也說明了“證成”與“說服”之間的差異。一方面，證成主要處理推理的“有效性”（validity），但這種有效性僅能以“全有或全無”的方式呈現，忽視了對論證有效性程度的考察。而論證的說服功能則以“充分性”（sufficiency）取代“有效性”之評價標準，這一標準兼顧論證的情境化特徵，即便某一論證在特定場景中是無效的，但仍有可能滿足其他領域的充分性標準。充分性也意味著不同論證對結論的支持存在程度之別。例如，有些法律論證僅能排除合理懷疑，而有些則可完全駁倒對方的論證。沃爾頓認為，“充分性”的標準顯然不是為了證明一方的正確性，而是對雙方論點的說服力進行評估，因此，論證“充分性”標準的運用顯然應當以語用為基礎。^[18]而相較於形式邏輯，非形式邏輯的優勢便在於，通過對語用維度的關注，使其得以實現論證的說服功能。

當然，在既往的研究中，也有一種聲音認為形式邏輯在某種程度上可以實現與非形式邏輯相同的效果。例如，如果對演繹推理的前提進行修改，將限制性條件加入到規則的構成要件當中，那麼同樣可以應對法律論證中的例外與規則衝突現象，進而實現與非形式邏輯近似的反駁功能。^[19]不

[16] See David Hitchcock, *The Practice of Argumentative Discussion, Argumentation*, 2002, p.287–298.

[17] 沃爾頓在其文中進一步豐富了論辯層的內容：（1）這一過程應當允許質疑，也即允許對方提出相反意見。現實的法庭辯論亦是如此，法官不可能僅聽取一方之言就得出判斷；（2）能夠標示證明責任的。例如“誰主張誰舉證”的法律證明責任，正方需為自己的主張承擔證明責任，如果反方提出質疑，那麼證明責任便隨之轉移；（3）這一過程的論辯屬性允許論辯主體撤回其結論，這意味着可以還原訴訟過程中的撤訴及對對方論證的認可。see Douglas N. Walton, *Defeasible Reasoning and Informal Fallacies*, Synthese, 2011, p.377–407.

[18] See Douglas Walton, *Informal Logic: A Pragmatic Approach*, Cambridge University Press, 2008, pp.1–3.

[19] See Peng-Hsiang Wang, *Defeasibility in der juristischen Begründung*, Baden-Baden: Nomos, 2004, S.143–173, 轉引自雷磊：《人工智能時代法律推理的基本模式——基於可廢止邏輯的刻畫》，載《比較法研究》2022年第1期，

過，尚且不說形式邏輯如何在缺乏語用維度的情況下實現法律論證的說服效果，即便在反駁功能上，這種方式亦是存在問題的。因為法律所應用的世界之開放性與不可預見性本質，使得我們無法在事前就羅列出所有潛在的例外情況。此外，就人工智能論證模型的開發目的而言，理想的情況是人工智能能夠在全樣本數據庫的基礎上實現自動裁斷，非形式邏輯的可廢止性特徵顯然更合乎這一目標。

三、基於非形式邏輯的智能論證模型

早在上世紀五十年代，圖靈（Alan M. Turing）提出了“機器會思維嗎”這一著名哲學問題，並嘗試通過一個被稱之為“模仿遊戲”（the Imitation Game）的思想實驗進行驗證。實際上，在邏輯學家看來，論證推理才是思維形式的核心，^[20]因此“機器是否具備思維”這一問題可以歸結為“機器是否會進行論證推理”。如此一來，法律人工智能的論證功能實現就聚焦到了法律論證的建模問題上。在本文所提出的論證模型中，非形式邏輯可以通過兩個核心的“組件”補充實現法律人工智能的論證功能：其一，是基於可廢止性所建立的、容納起應主體間相互反駁的“論辯結構”；其二，是基於語用維度、從聽眾視角出發，圍繞爭議焦點所建立的論證“可接受性”框架。二者的有效互動不僅能夠完成法律論證的證成環節，而且能夠對法律人工智能的論辯與說服功能進行補充。

（一）規則的邏輯形式化與“例外”的表達

需要明確的是，非形式邏輯並非是沒有“形式”的形式邏輯。如果我們將“形式”看作是受到規範、標準或程序的支配，那麼非形式邏輯本身也是“形式”的。非形式邏輯並不排斥邏輯語言的必要，而是拒斥將邏輯形式理解為論證結構的關鍵，並將形式有效性當作評價論證好壞的唯一標準。^[21]因此，對基於非形式邏輯的智能論證模式進行建模同樣可將規則及其例外進行邏輯形式化處理。

與一般的社會規範不同，法律規則所展現出的往往是一種條件式結構，也就是滿足特定的構成要件將會觸發某一法律後果，其中法定的構成要件作為結構的前件，法律後果作為後件。完美的條件式結構可以被形式化為：

$$E \leftarrow A_1 \wedge A_2 \wedge A_3 \dots \wedge A_n$$

在建模的層面其可以簡單表達為“如果……那麼……”的形式，E 代表“那麼”之後的法律後果，A 代表諸個構成要件。雖然法官在進行法律推理必須時刻遵循這一“法律 - 實事 - 結論”的流程，但是完美的條件式結構並不能複刻出法官在真實審判程序中的思維活動，規則的例外與特殊情況在實踐中往往在所難免。那麼，如何在完美的條件結構中加上“不完美”的成分？紐曼（Neumann Jr.）認為每一條法律規則必須包含“一組可被檢測的要素”，並且允許部分規則可以涵蓋一個或一個以上的例外，當這些例外情況出現時，即便所有的要素均被滿足，結論仍會被擊敗。^[22]沙托爾（Sartor. G）進一步將其具象化，認為法律規則構成要件分為兩個部分：“有待證明的要素”（elements to be proved）和“未被駁倒的要素”（elements not to be refuted）。“有待證

第15頁。

[20] 參見熊明輝：《法律人工智能的推理建模路徑》，載《求是學刊》2020年第6期，第91頁。

[21] 參見武宏志、周建武、唐堅：《非形式邏輯導論》（上卷），人民出版社2009年版，第135頁。

[22] See Richard K. Neumann, *Legal Reasoning and Legal Writing: Structure, Strategy, and Style*, Wolters Kluwer Law and Business, 2013, p.10.

明的要素”即需要當事人舉證進行確定，當一條規則中的此類要素皆被證實，並且“未被駁倒的要素”未被擊敗時，法官就可以確定法律規則的構成要件並得出結論。^[23]因此，規則的條件式可以被修改為：

$$E \leftarrow A_1 \wedge A_2 \wedge \langle A_3 \rangle \cdots \langle A_n \rangle^{[24]}$$

其中， A_1 、 A_2 作為“有待證明的要素”， $\langle A_3 \rangle$ 、 $\langle A_n \rangle$ 作為“未被駁倒的要素”。如此一來，將推理的規則前件劃分為以上兩種要素，便確定了不同要素在案件結果證立過程中的角色地位。

試舉一個例子來說明。《中華人民共和國民法典》（以下簡稱《民法典》）第 1245 條規定：“飼養的動物造成他人損害的，動物飼養人或者管理人應當承擔侵權責任；但是，能夠證明損害是因被侵權人故意或者重大過失造成的，可以不承擔或者減輕責任。”在此條規則中，動物的飼養人和管理人承擔無過錯責任，責任的阻斷條件需要由飼養人舉證。那麼綜合前兩種要素，該條可以表述為：

x 應當承擔侵權責任

IF 動物 z 由 x 飼養或者進行管理；

AND 動物 z 實施了行為 f

AND 行為 f 侵害了 y 的民事權益並造成了損害

AND \langle 被侵害人 y 不具有故意或重大過失 \rangle

在這當中，動物造成損害的具體事實作為“有待證明的要件”，需要由被侵害人給予證實， \langle 被侵害人 y 不具有故意或重大過失 \rangle 屬於規則前件中確立飼養人責任的“未被駁倒的要素”。這表明，在飼養動物致人損害的案件中，動物的飼養人和管理人通常需要承擔責任，除非其能夠證明被侵害人有故意或者重大過失。比如 y 出於好玩特意去挑逗 x 所飼養的狗，則此時該條要件將會被擊敗，進而影響意圖證立的結果。

基於非形式邏輯的法律論證模型需要一定程度上還原庭審中起應雙方主體的交鋒狀態，因此模型的構建需要建立在“衝突”的觀念基礎上。上述規則與規則的例外情況是法律推理過程中“立與破”的運行依據，規則的例外一定程度上能夠阻卻特定法律後果的獲得。在成文法中，規則的例外一般具有兩種形式：一種是例外情況本身就蘊含在法條之中。比如《民法典》第 985 條：“得利人沒有法律依據取得不當利益的，受損失的人可以請求得利人返還取得的利益，但是有下列情形之一的除外：……”。另一種是專門的例外條款，比如民法典有關贈與人撤銷權的規定。

我們同樣可以借助“未被駁倒的要素”將《民法典》第 1245 條的例外情況展示出來。該條規則中， \langle 被侵害人 y 不具有故意或重大過失 \rangle 作為證立飼養人責任的規則要件，實際上已經將其例外情況的否定集納入到了構成要件之中，也就是如果規則之例外表示為：

e: $\neg E \leftarrow A^e$

e: x 不應當承擔侵權責任

IF 被侵害人 y 具有故意或重大過失 (y 有意挑逗 z)

那麼上文的規則結構即為：

$$E \leftarrow A_1 \wedge A_2 \cdots (\neg A^e)^{[25]}$$

其中，規則之例外 (e) 承擔着對規則結論 (E) 的否定作用，規則中的“未被駁倒的要素”通

[23] 參見雷磊：《人工智能時代法律推理的基本模式——基於可廢止邏輯的刻畫》，載《比較法研究》2022年第1期，第16頁。

[24] See Giovanni Sartor, *The structure of norm conditions and nonmonotonic reasoning in law*, In Proceedings of the 3rd International Conference on Artificial Intelligence and Law, 1991, p.155.

[25] Ibid., p.158.

過例外情況的否定集 ($\neg A^e$) 進行表示。在智能論證模型的運行工作上，通過非形式邏輯實現的這種規則與例外的邏輯關係可以幫助我們建立起法庭論辯的對抗情景，而這正是傳統形式邏輯所無法達到的。

(二) 智能論證模型的論辯結構

在整體的模型建構中，我們希望達到的是通過多重例外（例外的例外）將規則中固有的“未被駁倒的要素”剔除，進而形成案件的“唯一正解”，或者僅留下爭議的焦點。因此可以構建起能夠展現論辯層級，並且反映出法律論證之後果不斷變化的論辯結構（圖 1）：首先，將各方所持的規範與事實前提吸收到論據 A 當中，當且僅當證明規則 ($r: \text{if } q \text{ then } p$) 的要件均在事實的佐證下得以滿足，A 論據才能夠推出 p 結論。其次，起應雙方能夠在此框架中實現不同論辯層級的對抗關係。在論辯 D_1 層面，起訴方的 A_1 論據能夠支持結果 C_1 的呈現；在論辯的 D_2 層面，應訴方所提出的 A_2 論據可以引出一階例外 e_1 ，進而破壞掉起訴方 A_1 對結果 C_1 的證成，也即案件 C_1 結果並不成立，引出結果 C_2 ；在論辯的 D_3 層面，起訴方所提出 A_3 論據同樣可以引出二階例外 e_{e1} ，攻擊應訴方的 A_2 論據進而恢復 A_1 論據對結果 C_1 的支持。以此往復，直至一方無法再圍繞爭議點提出與對方相反論據，或者至其論據無法再推翻對方論據。在這一過程當中，各方論據所包含的規範及事實促使了論辯雙方“攻防”角色的不斷替換，也進而使得法律結果隨之變換。

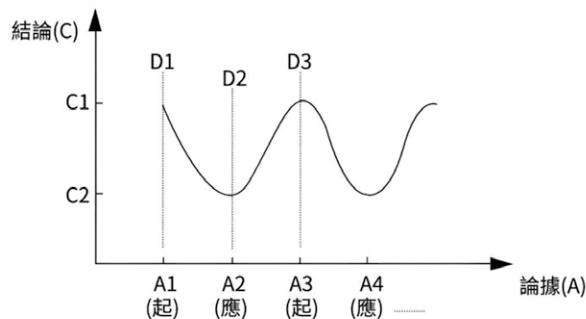


圖 1 智能論證模型的論辯結構

引入一個複雜情景進行說明：甲育有一子乙。丙為甲的朋友，在甲身患重病期間悉心照料甲的生活起居，甲遂立遺贈撫養協議將自己的部分遺產遺贈給丙。在甲病逝後，乙、丙分別向法院起訴，主張各自的法定繼承權、受遺贈權。其中，乙援引《民法典》第 1127 條^[26]，證明自己系甲的法定繼承人，因此有權繼承遺產。此條規則可以表示為：

- $r_1:$ 乙可以作為甲的法定繼承人繼承甲的部分遺產
 - IF 甲已逝世； (a)
 - AND 甲有合法有效且可執行的遺產； (b)
 - AND 乙係甲的兒子； (c)

丙援引《民法典》第 1158 條：“自然人可以與繼承人以外的組織或者個人簽訂遺贈扶養協議。按照協議，該組織或者個人承擔該自然人生養死葬的義務，享有受遺贈的權利。”該條規則可以表示為：

- $r_2:$ 丙可以作為受遺贈人取得甲的部分遺產
 - IF 甲已逝世； (a)

[26] 《中華人民共和國民法典》第1127條第1款第（一）項：“遺產按照下列順序繼承：（一）第一順序：配偶、子女、父母”。

AND 甲有合法有效且可執行的遺產； (b)

AND 丙盡到了生前扶養、死後喪葬甲的義務； (d)

AND 甲、丙之間有合法有效的遺贈扶養協議； (g)

假設，上述的所有事實均已被證實，那麼乙、丙的主張就可以被法院所允許。此時，乙的論據可以表示為 $A_1 (r_1, a, b, c)$ ，丙的論據可以表示為 $A_2 (r_2, a, b, d, g)$ 。繼續假定兩條來自雙方提供的事實與規則（例外）：

f_1 (丙的新論據)：乙曾經嚴重虐待過其父親甲

f_2 (乙的新論據)：丙曾經為了爭奪遺產試圖殺害乙，以此來增加自己所能夠享有的利益。

所適用的規則皆為《民法典》第 1125 條第 1 款有關繼承人與受遺贈人喪失繼承權與受遺贈權的規定，^[27] 其可以表示為：

e_1 : 乙無權繼承甲的部分遺產

IF 乙曾經嚴重虐待過甲；

e_2 : 丙無權接受甲的遺贈

IF 丙曾經為爭奪遺產試圖殺害乙

此時，雙方都提出了一條針對乙、丙所援引規則的例外情況，也即當乙、丙實施了《民法典》1125 第 1 款規定中的行為時，原本所具有的繼承權與受遺贈權喪失。如果 f_1, f_2 均已證實，這便構成了雙方用以攻擊 A_1, A_2 論據的反向論據 $A_3 (e_1, f_1)$ 與 $A_4 (e_2, f_2)$ ，根據例外對規則的可廢止性，也即 $e_1 > r_1, e_2 > r_2$ ，原本的 A_1, A_2 論據將會被新論據所擊敗，法律推理的後果也隨之變化。為了展現出論辯的多重性，我們可以在此論辯層級上再加入一個反駁項，繼續引入新的事實及依據。乙可以提出事實 f_3 ：甲生前曾通過書面形式對乙表示寬恕。^[28] 在此情境中，此條例外的例外可以表示為：

e_{e1} : 乙可以作為甲的法定繼承人繼承甲的部分遺產

IF 乙確有悔改表現

AND 甲生前曾通過書面形式對乙表示寬恕

由此，乙新生成的論據 $A_4 (e_{e1}, f_3)$ 便可以作為 $A_3 (e_1, f_1)$ 論據的反向論據將其擊敗，以此恢復 $A_1 (r_1, a, b, c)$ 論據的效力，進而回歸到原本的案件結論上。

不難看出，基於非形式邏輯的論辯結構能夠破除既有研究對主體的回避，還原起應雙方的主體地位，並且很好地處理了規則與例外情況的關係。在傳統的形式邏輯模型中，規則與例外的處理往往是將規則之例外統一涵蓋在大前提當中，多以“but”邏輯詞相連接。這種方式一方面無法生動還原法庭論辯的互動狀態，使整個法律人工智能陷入單一、呆板的境地。另一方面，且不說能否用一個指令完整涵蓋所能涉及案件的例外情況，在法律適用的疑難情況中，例外情況往往無法預先設定，只能在雙方的論辯過程中逐漸生成。此外，在上述例子中，各方論據之間的爭鋒狀態需要建立在規範與事實的有效結合上，這一定程度也解決了審判程序中的舉證責任分配。雖然系統並沒有明確展現證明責任的分配，但其可以在特定的規則、行動集之間的互動過程中將“誰主張誰舉證”這

[27] 《中華人民共和國民法典》第1125條第1款：“繼承人有下列行為之一的，喪失繼承權：（一）故意殺害被繼承人；（二）為爭奪遺產而殺害其他繼承人；（三）遺棄被繼承人，或者虐待被繼承人情節嚴重；（四）偽造、篡改、隱匿或者銷毀遺囑，情節嚴重；（五）以欺詐、脅迫手段迫使或者妨礙被繼承人設立、變更或者撤回遺囑，情節嚴重。

[28] 《中華人民共和國民法典》第1125條第2款：“繼承人有前款第三項至第五項行為，確有悔改表現，被繼承人表示寬恕或者事後在遺囑中將其列為繼承人的，該繼承人不喪失繼承權。”

一原則進行內化。

(三) 智能論證模型的“可接受性”框架

就模型的構建而言，僅僅恢復起應雙方的主體地位仍舊不夠。法律論證的論辯結構可以描述控辯雙方之間論證的反駁與被反駁、擊敗與被擊敗關係，但是論據之間的相互作用同樣需要明晰。也就是說，各方的論證在什麼情況下才能在論辯中“勝出”？傳統形式邏輯將論證形式的有效性作為判斷的標準，這通常難以適應複雜的法律環境。對話及修辭的理論讓我們關注到“目標聽眾”在這一過程中的重要地位，應當通過目標聽眾對論證的可接受性來評價論證的“優劣”。因此，文章接下來將率先考察法律論證環節的目標聽眾，進而對論證模型的“可接受性”框架進行描述。

在訴訟環節中，起訴方、應訴方、與審判方的目標聽眾並不完全相同。起訴方與應訴方在訴訟過程中的目標聽眾包含自己、對方及審判方。說服自身是起、應雙方論證的第一層目的。起訴方作為訴訟博弈的主動方，是否進入到訴訟階段，主動權完全掌握在起方。因此，對於民事訴訟來說，理性的起訴方會在進入訴訟程序之前認真權衡訴訟為自己帶來的收益，只有說服自身認為能夠從此獲益才會提起訴訟。同理，應訴方要想使自己的論證能夠成功反駁對方，首先也需要說服自身，但是是否進入到訴訟程序其並不能自行決定。說服對方接受自己的論證一般在進入訴訟程序之前，進入到訴訟博弈階段，說服對方就不那麼重要了。此時，二者會圍繞案件的爭議焦點據理力爭，無法實現一方被說服的局面。雙方之所以能夠接受最後的結果往往並不是因為對方給出了一個好的論證，而是因為接受程序正義的法律原則。^[29]這一階段，說服審判方才是雙方論證的主要目的。

審判方的目標聽眾相較於前二者較為特殊。根據佩雷爾曼（Perelman）和蒂特卡（Lucie Olbrechts-Tyteca）對目標聽眾的區分，目標聽眾可以包括一類具體的人群，也可以被認為是所有的理性人。前者是論證的特定聽眾，後者則是普遍聽眾。^[30]審方論證的目標聽眾也可以進行如此區分，其具體聽眾由判決書所直接影響的人群構成，包括審方自己、起訴方和應訴方。普遍聽眾即上級法院、法學家、普羅大眾等所有理性人組成的法律共同體。

為了引發目標聽眾的共鳴，論證的說服功能決定了其評價需要引入修辭學的思想，非形式邏輯所關注的語用維度又恰當地將這一思想表達出來。不論是古典修辭學還是佩雷爾曼所提出的新修辭學，都是一門研究如何說服目標聽眾的藝術。新修辭學更是進一步將論證視作說服聽眾的手段，認為應當將聽眾對論證的可接受性作為評價標準，而非單純的形式有效。法律人工智能的論證模型構建可以通過一套程序性方法，允許上述聽眾在模型中尋求有關論證的解釋，從而將論證的“可接受性”表現出來。

論證的“可接受性”評價過程可以分為三個部分（圖2）。第一，論辯結構中未被起應雙方所接受的爭議焦點。如前所述，起訴方與應訴方作為互相說服的論證主體，真正能夠接受對方論證的時間主要在訴前的階段。這一經過可能是雙方訴前的和解，也可能是各自對論證可接受性的內心推演，而未被雙方所互相接納的論證將形成爭議的焦點。

第二，訴訟過程中的論證主體向聽眾進行解釋論證，這一過程又可區分為兩個層面。一方面，起應雙方需要圍繞爭議焦點進行解釋論證，目標聽眾是法官。此時，人工智能作為提問者可以向論證的雙方提出解釋要求，起應主體需要對論證中的法律規範及案件事實做出進一步解釋。另一方面，法官需要對其所做出的法律決策面向起應雙方及普遍聽眾做出解釋。此時，人工智能則作為解釋者對其生成的判決預測、刑期預測及法律文書做出解釋。第三，聽眾對解釋的檢驗階段。不論是旨在說服審方的起應主體，還是要對其法律決策做出廣泛解釋的審方，在對其論證進行解釋之後，

[29] 參見熊明輝：《訴訟論證——訴訟博弈的邏輯分析》，中國政法大學出版社2012年版，第188頁。

[30] See Eveline T.Feteris, *Fundamentals of Legal Argumentation*, Springer, 2017, p.45-46.

都需要各自的聽眾來檢驗解釋是否通過，進而反映論證的可接受性。如果解釋通過，那麼解釋者便成功使聽眾接受其立場。反之，則意味著解釋沒有通過。此時，聽眾可選擇進入新的一輪解釋論證中，即允許解釋者從其他角度再次進行解釋，直到解釋通過，或者直接選擇不接受解釋者的論證。

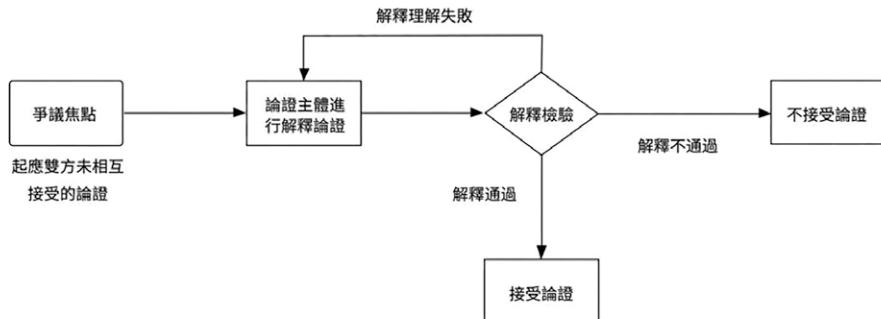


圖 2 智能論證模型的“可接受性”框架

接續前文的事例。乙與丙在爭議焦點產生之前會經歷一個互相說服的階段。從論辯結構來看，乙提出的論據 1 能夠擊敗丙所提出的論據 2 主要有兩個原因：一種是論據 1 要挑戰論據 2 中規則 r 所產生的結論，也即對規則所產生的後果進行否定，這種情況下需要比較二者之間所包含的子論據（規則與例外）之間的強度。一般來說，我國成文法中所規定的例外性條款普遍優先於規則本身，也即 $e > r$ 。另一種是論據 1 需要阻斷規則 r 與結論之間的聯繫，證明待適用的規則並不合適，而不是攻擊後果本身。但衝突的結果還有一種情況是各論據之間的強弱關係很難比較，進而形成案件的爭論焦點，控辯雙方的說服目標也隨之發生轉變。不妨將上文事例再做出一些調整：

丙仍舊援用《民法典》第 1125 條第 2 款第（三）項作為依據，只不過在事實方面將原本的 “ f_1 : 乙曾經嚴重虐待過其父親甲” 修改為 f'_1 ：乙曾毆打其父親甲，並且甲在疾病影響下心理素質堪憂，每次遭遇此事都會寢食難安導致體重驟減，也一定程度上影響到了身體的恢復。^[31] 此時，如何權衡 f'_1 與 e_1 所形成的新論據 $A_6 (e_1, f'_1)$ ，與最初的論據之間的擊敗關係？在實際的審判過程中，法官可能會通過虐待行為的手段、次數、實施的時間及後果來判斷是否應當將這種情況納入到“情節嚴重”當中。審方作為起應主體的目標聽眾可以請求其對論證作出進一步的解釋：

開始階段：

解釋聽眾 [請求解釋]：請解釋為何認定乙喪失繼承權？

解釋者丙 [宣告]：乙曾經虐待過其父親甲，且情節嚴重。

對話階段：

解釋聽眾 [質疑]：乙採取何種手段虐待其父親甲，並且施虐的次數及時長是多少？

解釋者丙 [解釋]：乙通過拳打腳踢的方式，每日虐待其父親甲數次，時間大約持續一小時之久。

解釋聽眾 [請求解釋]：通過什麼證據進行佐證？

解釋者丙 [解釋]：家中的監控器及家政服務人員可以證明（提交證據）。

結束階段：

解釋聽眾 [接受]：我接受你的論證。

當然，現實司法實踐中的論證活動並不會如此簡單，案件事實的複雜性以及規則適用的彈性，

[31] 做出此調整的目的在於，將原本事實中的“嚴重”情況模糊化，這一定程度上將會影響法官判斷是否符合法條中的“情節嚴重”。

往往會使解釋論證的結構呈現出更為複雜和多層次的面貌。不過，整個模型對論據的評價基礎卻是統一的，也就是需要通過以語用維度為導向的“充分性”標準檢驗論證的成功與否。此外，需要強調的是，在這一案例中，“可接受性”框架被適用於評價起應主體的論證，但論證活動仍舊需要通過論辯結構展開。比如，目標聽眾也可能會接受來自反方的解釋論證，更有力的反駁——例如，乙已獲得其父親甲的生前諒解——可以改變原本論證的可接受性狀態，進而否定丙的解釋論證。

在技術配套問題上，人工智能對非形式邏輯的論證與評估理論在實踐中也得到了驗證。典型的系統代表是人工智能辯手系統。相關的研究團隊證明了論證實踐以及對論證的評估能夠被非形式邏輯清晰表達，並發現理論與自然論證對論證實踐的評價是相同的。^[32]這類人工智能辯手系統從一般意義上區分了“論證品質”的3個維度：邏輯、修辭和合理性。邏輯維度的評價包含前提的可接受性、前提與結論的相關性以及前提對結論的支持程度；修辭維度的重點關注論據的可信度、情感訴求、清晰度和適當性等；合理性維度包含聽眾對論證的接受程度、論證對最終結論的支持程度以及論證能夠充分反駁可預見的反對意見的屬性。^[33]

四、智能論證模型的其他考量因素

不可否認的是，非形式邏輯與人工智能技術理性的有機耦合雖然能夠作為智能建模的新動力，而且也取得了相當大的技術進展。但這種以非形式邏輯驅動的人工智能建模方式仍然是一種情境化的應用。拋卻論證邏輯本身，法律論證實踐仍舊存在諸多主體性色彩的因素。因此，如果要對模型的建構作出更為全面的考慮，那麼至少還應包含對以下幾種因素的考量：

（一）完善的證據論證

在論辯形式當中，法律的適用、案件事實與案件結論應當是相互交織地發展並完善。法律適用需要考慮的是，一條待適用的規範能否可以將待決案件涵蓋在其要件之下；證據推理則是在已取得的事實證據基礎上重新構建法律事實，以判斷案件事實是否符合各種要件，然後進行法律論證。本文在模型的建構中雖未重點突出證據論證，但考慮到“法律+事實=案件結論”的推理結構，法律論證的運行過程理應展現出一種規範與事實交互前進的姿態，動態地剔除和肯定各類信息，從而引出合理的結論。因此，智能論證模型的完整構建應當將證據推理也納入到研究範疇。

一般來說，證據推理模型的構建主要存在三種進路：基於論證進路、基於故事進路以及概率推理。^[34]基於論證進路主要源自於威哥摩爾（John Henry Wigmore）的證據理論，其理論以事實的證明力作為基礎，研究個體事實的證明力在證據鏈條中如何分配的問題。據其理論，案件中各方所提出的證據群都可以以相互連接的推論網進行表達，並將法律證據建立在可以承載證明力的似真推論的基礎上，這些推論隨着案件的進展將有可能被新引入的證據所擊敗。^[35]基於故事進路的證據推理源自於法律心理學，班尼特（W. Lance Bennett）等人認為刑事證據的運用應當是借其構建可能案情經過的假設性故事，並通過各種既定的標準選擇出最佳故事。之後的學者又在其基礎上作出了進一

[32] See Henning Wachsmuth, et al., *Argumentation Quality Assessment: Theory vs. Practice*, in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, p.250–255.

[33] See Henning Wachsmuth, et al., *Computational Argumentation Quality Assessment in Natural Language*, in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017, p.182.

[34] 參見杜文靜、蔡會明：《法律論證的人工智能模型》，載《上海政法學院學報（法治論叢）》2019年第1期，第56頁。

[35] [加]道格拉斯·沃爾頓：《論辯與證據理論》，王建芳、童海浩譯，載陳金釗、謝暉主編：《法律方法》2015年第1期，山東人民出版社2015年版，第25頁。

步的延伸，提出了如何運用故事片段並分析已有證據。^[36] 基於論證的進路與基於故事進路都是在對案件事實進行定性的分析，概率推理則是目前最貼合法律人工智能模型構建的實踐路徑，其採用定量分析的方法，通過貝葉斯信念網絡評估證據與待證事實之間的契合度。例如，運用目的導向性的信念網絡對複雜案件進行建模，或者作為概率專家系統分析並評價一系列具體的法庭證據的信息價值。^[37]

（二）必要的價值判斷

基於非形式邏輯所刻畫的智能論證模型能夠從修辭理論的視角重新界定“好論證”的評價標準，避免了形式邏輯的機械性、封閉性評價。但是價值判斷仍舊是人工智能法律論證的核心難題。因為法律人並非將法律規則視為片面的法律表述，而是用以追求特定目的的手段，這終究會回歸到各種價值利益之間的衡量。相較於人類思維對論據的經驗性理解，人工智能無法洞悉不同論據的真正“意義”。同時，不同論據之間的強弱關係，論據對特定結論的“支持”或“反對”關係往往較為複雜，作為裁判依據的各項規範代表了不同的價值取向，旨在說服的目標聽眾對同一事件的法律判斷、道德判斷本身就存在不一致。比如，在“瀘州遺贈案”當中，學界對本案的判決持兩種針鋒相對的觀點。一種觀點認為應當按照法律的規定，黃某的遺囑有效，因此應按遺囑內容分配給張某部分財產。而另一種觀點則認為黃某與張某並不是法定的夫妻關係，而是為大眾所不能接受的“非法同居”關係，其贈予行為有悖公序良俗，因此應當認定遺囑無效。顯然，本案涉及到法治與道德這兩種價值訴求的矛盾，這是法官在處理疑難案件時經常會面臨的不可通約性難題。

不少學者在這一難題上嘗試過理論突破。日本學者平田勇人自 2006 年開始在諸多數學家的幫助下研究如何根據法律推理的主體、語境以及價值判斷的功能之間的關係確立價值函數，從而完成智能價值植入。^[38] 其基本主張是將法律的價值判斷轉化為價值函數，將不同類型審判所涉及的基本價值作為價值函數的演算符，比如民法中的公平原則、平等原則等。這當中每一個演算符都會根據不同的價值評價標準進行獨立運算，最後將各評價值進行綜合評價。如若存在價值評判標準的衝突或者其他圍繞評價標準本身的問題，那麼可以在論辯模型中展開以達成適用價值的妥協為目的的商談活動，進而達成價值共識。實際上，對價值進行賦值仍舊是以自動化的方式將預設的輸入值與一組可能的輸出結果相關聯，而非復現法律論證。如此以來，法律論證中的邏輯關係便成為了參數間相關性的比較，論證的可靠性很大程度上取決於數據的質量與算法的設計。這是目前弱人工智能應用開發中的“智識瓶頸”。不過，雖然不能使人工智能實現真正意義上的法律論證，但同人類思維模式的建立過程相似，在缺乏主觀判斷的情況下仍舊可以通過海量歷史數據的學習積累，擴大人工智能預設輸入值的範圍，提高輸入數據的質量，進而無限接近於嚴格意義上的法律邏輯思維。

（三）人為的干預與更新

鑑於目前人工智能在還原法律論證思維上的困境，智能論證模型從弱到強的技術轉型階段免不了人為的干預及更新。一如上文所討論的那樣，在事實認定領域中，“自由心證”賦予了裁判者根據內心確信評價證據證明力的權力，這種裁判的零散性與不確定性使得人工智能短期內無法完全勝任裁判者一職。而對於價值判斷的運用，人工智能更是力所不逮，即便能夠在人為標註價值權重的幫助下解決部分簡單案件，但在面臨重大疑難案件時仍舊需要人為的干預。

[36] See Floris J.Bex, *Arguments, Stories and Criminal Evidence*, Springer, 2011, p.2-3.

[37] See Amanda B.Hepler, A.Philip Dawid & Valentina Leucari, *Object-oriented Graphical Representations of Complex Patterns of Evidence*, Law Probability and Risk, 2007, p.2-22.

[38] 參見季衛東：《人工智能時代的法律議論》，載《法學研究》2019年第6期，第42-43頁。

此外，考慮到基於非形式邏輯的論證系統的語用性質，知識庫的語義更新與人為標示也是不可或缺的。如上文所述，語用是一個在對話情境中進行話語明示的過程。話語要在具體語境中發揮作用，並不能僅僅依靠語句中概念的字面意思，更多的應該推導出語句在特定情境下的意蘊。在我國，作為成文法規範之對象的現實生活往往處在千變萬化之中，新出現的社會情況經常不必然被已有的規範語句所包含，仍舊用過去法律語句的含義來概括新出現的事實可能會導致語詞的無效。以“許霆案”為例，在當時的時代背景之下，“利用 ATM 機的故障獲取錢財”並不在當時 1997 年刑法的語境之內，因為這絕對不是當時立法者的所持立場能夠涉及的。因此，系統之中應建立完善的更新機制來幫助法律人工智能做出更合理的判斷。

總而言之，人工智能目前的諸多技術應用實現了非形式邏輯在法律論證領域的工具論價值，也檢驗了其理論成果的實踐效力。但非形式邏輯驅動的法律人工智能論證系統仍舊無法回避論證中的主體性特點。因此，在發揮技術層面的客觀性、穩定性的同時，我們也要在實踐層面正視經驗於論證活動中的價值，清醒認識到法律人工智能的優勢與限度。

結語

在某種程度上，為實現法律論證的反駁與說服功能，法律論證應當是一種建立在對話基礎上的辯證推理，同時也是在特定場景下對特定聽眾的一種說服。基於形式邏輯所刻畫的法律人工智能因其“機械”的演繹形式而無法還原法律論證的自然屬性。因此，人工智能論證模型的建構需要接受非形式邏輯的路徑轉向。非形式邏輯否定形式邏輯“零主體”、“無目的”的基本特徵，為刻畫庭審主體的論辯活動留有空間，從而更好展現論證的反駁功能。同時，非形式邏輯考慮到了論證的語用維度，將論證的評價標準從“形式有效”替換為聽眾的可接受性，進而補充了法律論證的說服功能。反映在模型建構上，基於非形式邏輯的智能論證模型能夠通過論辯結構，模擬起應主體之間的對抗狀態，並將聽眾納入到論證的評價體系當中，通過“可接受性”框架還原法律論證主體為說服目標聽眾所進行的解釋性活動。除此之外，法律論證還需包含對事實因素的考量，應將證據推理的形式化納入到模型設計當中。同時，人工智能現階段的技術發展樣態仍舊無法妥善應對“自由心證”、價值衡量、語用確認等主觀判斷環節，此類系統運行的特定階段仍需以人機協同的方式進行。

Abstract: In general, argumentation serves three primary functions: justification, rebuttal, and persuasion. Traditional AI argumentation systems that rely on formal logic often overlook the plausibility, dialectical nature, and pragmatic aspects of legal reasoning, limiting their focus primarily to justification. Non-formal logic provides a new method for modeling legal argumentation in AI, addressing the shortcomings of formal approaches. This paper seeks to incorporate non-formal logic into the construction of legal AI argumentation models. By focusing on the concept of “defeasibility”, we will explore rebuttal in AI legal reasoning from both a dialogical and argumentative perspective. Through the pragmatic dimension of non-formal logic, we will further develop a framework of “acceptability”, which highlights how to persuade the target audience as a crucial standard for evaluating the adequacy of arguments. Additionally, the evidentiary reasoning and the presence of subjective factors in value judgments remind us of the inherent limitations of computational models, indicating that appropriate human intervention is sometimes very important.

Key words: Legal Artificial Intelligence; Function of Argumentation; Non-formal Logic; Argument Model

(責任編輯：張雨虹)