逻辑学研究　2023 年第 4 期，40–54
文章编号：1674-3202(2023)-04-0040-15

本土社会价值如何融入人工智能论辩系统
——结合广义论证理论与基于价值的实践推理

余喆

摘　要：现实世界中的论辩必然发生在某种社会文化背景和动态变化的语境之下，其中涉及特定的社会规范和社会价值。对于实践推理决策而言，参与者之间的意见分歧往往不仅在于常识和信念，更在于对社会价值的不同优先排序。本文结合考虑语境的形式论辩理论，讨论如何将本土社会价值融入人工智能论辩系统，使得持有不同价值观的参与者之间能够建立有意义的共识。在对论证实例进行分析的基础上，可以将相关建模方法扩展到基于价值的实践推理应用中，探索如何结合形式论辩、广义论证理论等跨学科研究成果，评估考虑社会价值的实践推理论证。

关键词：广义论证；论辩系统；实践推理；社会价值

中图分类号：B81　　**文献标识码：**A

1　引言

人工智能论辩系统可以视为旨在模拟基于论证/论辩的人类推理，并支持论证评价和推理决策的可计算模型。近二十多年间，形式论辩研究在该研究领域中备受关注，尤以人工智能界学者 Dung 所提出的抽象论辩理论（[10]），以及包含论证结构刻画的多种结构化论辩理论（[6, 11, 15, 18]）为代表。其中，抽象论辩理论作为一种非单调推理手段，致力于捕捉人们在诸如常识推理、实践推理等知识或信息不完全、不确定、不一致的情境下推理的基本特征，并基于论辩语义，对论辩框架中的论证进行评价和推理。这一机制主要可以实现论证评价的可计算化，但该理论仅将论证视为抽象的实心节点，并将论证之间的交互关系表达为一种简单的冲突（攻击）关系。结构化论辩理论则在这一关键环节的基础上，致力于完善从知识表示、论证构建、攻击关系识别，到（基于抽象论辩或是其他可计算方法的）论证评价和结论输出的过程。形式论辩系统由此可以完成一个从底层的具体知识表示，到顶层的抽象计算，最后得到符合理性的可辩护推理结论的基本流程。

收稿日期：2023-05-15
作者信息：余喆　　中山大学逻辑与认知研究所
中山大学哲学系
yuzh28@mail.sysu.edu.cn
基金项目：国家社会科学基金重大项目“汉译逻辑术语本土化与中国逻辑学话语体系建设研究”（21&ZD065）。

由此可见，目前常见的论辩系统要么主要从抽象的角度考虑论证及其交互，要么基于符号逻辑进行知识表示，定义论证构造、攻击关系等概念，构建结构化论辩框架或系统，并以抽象论辩理论等方法为计算评价手段。([5]) 从沟通以论辩为代表的人类推理和机器推理的目的来看，可以说，现有的形式论辩系统已经给出了一个看似可行的架构。然而，由于系统设计主要由抽象和形式的角度出发，论辩系统在面向具体应用领域时依然难以避免地存在着“落地”困难的窘境，众多国内外学者也先后指出了相关问题，例如，论辩系统设计中的合理性与建模人类推理时的自然性难以调和的问题（[7]），以及论辩系统的过度抽象化和脱离语境所导致的推理结果违背直觉的问题（[17, 26]）等。其中的一个关键点在于如何使论辩理论真正融入不同的社会文化情境，克服形式系统应用中水土不服的问题。

社会文化因素是目前论辩逻辑研究，尤其是形式论辩方面研究中受到忽视的部分，也是造成相关理论实际应用困难的关键因素之一。论辩所发生的具体语境中包含社会文化背景、社会规范、价值与伦理等一系列不同层面上的组成元素，要解决论辩理论和形式系统的可应用性问题，论辩情境中的关键因素不可缺席，而经过系统计算输出的结果也应当适应特定的社会文化认知。广义论证理论从包容多元文化的角度出发，提出了论证研究的本土化原则，即“某一文化中的论证只有在本文化中才能得描述和评价”，认为论证是“某一社会文化群体的成员，在语境下依据合乎其所属社会文化群体规范的规则生成语篇行动序列”。([26]) 该理论中尤其强调了论辩过程中的社会文化背景、社会规范与价值，以及语境敏感性等特点，所提炼出的关键因素正可以为上述难题的解决提供思路。同时也有学者分析指出，基于广义论证概念的逻辑观有助于真正从本土文化源流出发，还原传统文化的本来面目，挣脱西方文化主导逻辑观的禁锢，推动中国文化复兴。([24, 25])

当前，在以大数据和机器学习为核心人工智能技术取得令人瞩目的成果的同时，人工智能可解释、机器伦理，以及关于认知推理及其动态性的处理能力等方面的问题也引起了人们的担忧。形式论辩在这方面的优势在于：作为一种处理不一致情境下知识表示与推理的通用机制，易于结合偏好、权重、概率等决策因素，可以灵活应对推理中信息的动态变化，也可以基于逻辑与论证建立易为人们理解的人工智能可解释机制。([29]) 而另一个重要的研究课题，正如上文所指出的，在于人类需要利用人工智能系统处理他文化中的推理问题。为了实现这一目标，通常认为需要首先将不同文化中的表达与推理形式化。根据广义论证理论提出者鞠实儿的观点（[27]），这种形式化的表达与推理可能实际上异于他文化中的表达与推理，因而不是后者的如实表达。它的合理性判断标准应当在于是否有助于我们用人工智能系统解决他文化中的表达与推理问题。所以，在广义论证的视野下，形式化方法无疑具有工具的合理性，但形式语言的抽象描述、追求普遍形式表达的论证模式导致论辩系统忽视了论辩中的语境敏感性，以致相关理论凌空悬浮，难以指导生活。

在智慧化时代背景下，让论辩系统发挥出自身的优势，实现在论证挖掘技术、自然语言处理、可解释人工智能等新兴前沿重要领域中的本土化应用，需要人们扎根现实世界，突破学科壁垒。有鉴于此，本文旨在融合广义论证理论思想，基于形式论辩研究，从社会文化层面出发，重新审视人工智能论辩系统的构造理念，力求实现更好的学科交叉弥合，并推进化解论辩系统在可应用性上所面临的困难。根据这一思路，本文首先在第 2 节中提出一个考虑（与社会规范相关联的）价值偏好的论辩系统，讨论如何在人工智能论辩系统中结合社会规范与价值因素，并在处理不一致情境下的认知推理的同时，完成同一社会文化背景下多主体参与的实践推理决策；然后，在第 3 节给出与相关工作的比较；最后，本文第 4 节总结全文，并指出未来可供拓展的研究方向。

2 考虑价值偏好的论辩系统

根据广义论证理论，论辩是在动态开放语境下的一种社会互动。语言交流中的地域、历史、文化背景，参与者的知识信念系统、社会属性等都属于语境的组成部分。（[26]）因而，融入本土原则和语境原则的论辩系统应当既适用于处理关于知识和信念的认知推理，也适用于处理关于行动决策的实践推理。由于形式论辩系统本身的通用性特点，认知推理中的语境特征可以通过在论辩系统所构建的知识库上附加关于信念的元素的优先排序来体现，而实践推理则与论辩参与者的价值观有关，因此我们还需要在论辩系统中引入社会规范和社会价值的集合。

本节结合文献 [14, 15] 中给出的结构化论辩框架 *ASPIC*$^+$，并参考前期系列工作 [22, 23] 等中所提出的考虑语境的论辩框架，在形式论辩系统中融入社会规范和价值等实践推理元素，给出如下在某种语境下包含社会规范和相应的价值集合的形式论辩系统定义。由于价值因素在其中起到关键性作用，而包含知识库的论辩系统通常被称为一个论辩理论（Argumentation Theory，简称 *AT*），因此本文将下述定义中所提出的论辩理论简称为 *VAT*。

定义 2.1 (形式论辩理论)**.** 考虑社会规范和价值的论辩理论 *VAT* 可表示为一个元组 $(\mathcal{L}, \mathcal{K}, V, \mathcal{R}, n, val, \leqslant, \leqslant', P)$，其中

- $\mathcal{L}$ 是一种用于知识表示的逻辑语言，使之在否定关系（$\neg$）下闭合[1]；
- $\mathcal{K} \subseteq \mathcal{L}$ 表示一个知识库，使得 $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$；其中 $\mathcal{K}_n$ 是确定的公理性知识集合，$\mathcal{K}_p$ 是不确定的普通知识集合，二者之间不相交，即 $\mathcal{K}_n \cap \mathcal{K}_p = \emptyset$；
- $V = \{v_p, v_q, \ldots, v_z\}$ 表示一组价值的集合，下标“$p, q, \ldots, z$”各代表某种具体的社会价值；

[1] $\varphi = \neg\psi$ 或 $\psi = \neg\varphi$ 的情况可以记为 $\varphi = -\psi$。

- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d \cup N$ 是推理规则的集合，由三个不相交的集合 $\mathcal{R}_s$、$\mathcal{R}_d$ 和 N 组成，其中 $\mathcal{R}_s$ 和 $\mathcal{R}_d$ 分别表示硬性规则的集合和（除 N 中的元素之外的）可废止规则的集合，硬性规则的形式为 $\varphi_1, \ldots, \varphi_n \longrightarrow \varphi$（$\varphi_i, \varphi$ 都是 $\mathcal{L}$ 中的元素），$\mathcal{R}_d$ 中的可废止规则的形式为 $\varphi_1, \ldots, \varphi_n \Longrightarrow \varphi$；$N$ 表示一组社会规范的集合，形式为 $\varphi_1, \ldots, \varphi_n \overset{v_u}{\Longrightarrow} \varphi$，其中“$\Longrightarrow$”同样表示一种可废止的推理关系，$v_u$ 则表示与该规范相关联的价值“u”；
- n 是一个用于命名的偏函数，使得 $n : \mathcal{R}_d \cup N \to \mathcal{L}$，即为每条可废止规则指派一个唯一的名称；
- val 是从 N 到 V 的映射，对应 N 中的规范所关联的价值（即推理规则上标的“v_u”）；
- $\leqslant$ 和 $\leqslant'$ 分别表示 $\mathcal{K}_p$ 和 $\mathcal{R}_d$ 上的预序；
- $P = \{\lesssim_1, \lesssim_2, \ldots, \lesssim_n\}$ 是价值集合 V 上的预序的集合，其中的每个 $\lesssim_i \in P$ 代表一个论辩参与者（或者持方）对于价值的优先排序。

在上述定义中，$\mathcal{K}$ 中的元素可以作为论证构造的初始前提，$\mathcal{R}$ 中的元素是基于前提继续构造论证的规则，由此可以生成论辩系统中的论证集。而作为一种非单调推理手段，上述论辩理论中表示不确定性元素的集合有三个，分别是普通知识集[2] $\mathcal{K}_p$、可废止规则集 $\mathcal{R}_d$，以及规范集 N。按照本文对论辩系统的设定，仅使用 $\mathcal{K}_p$ 和 $\mathcal{R}_d$ 中的元素构造的论证表达的是纯粹关于知识和信念的认知推理，而如果论证构造中使用了规范，必定涉及某种与之相关联的价值，那么表明该论证有关行动决策，表达的是一个涉及实践推理的过程。需要说明的是，为了不使当前论辩理论过于复杂，本文假设了每个社会规范只关联一种主要的价值。事实上，根据实际应用的需要，也可以定义每个社会规范关联一组价值的集合。在上述三种集合之上，论辩参与者可以分别对其中的元素赋予不同的优先排序，即上述定义中的 $\leqslant$，$\leqslant'$ 和 $\lesssim_i$。由于本文主要考虑的是某种文化背景下的论辩（而非跨文化交流中的论辩），我们假设参与者在知识和信念方面已具有一定的共识。[3] 因此，定义 2.1 中主要藉由集合 P，着重体现同一社会文化语境下，不同持方的论辩参与者对于价值的不同排序。

在论证的结构中，参考 [14, 15]，需要至少标记几个关键的组成部分：设 A 为一个论证，Prem(A) 表示论证 A 的前提的集合，即 A 中所使用的所有 $\mathcal{K}$ 中元素的集合；Conc(A) 表示 A 的结论；Sub(A) 表示 A 的所有子论证的集合；DefRules(A) 表示 A 中所使用的所有普通可废止规则（即 $\mathcal{K}_d$ 中的规则）的集合；Norms(A) 表示 A 中所使用的所有社会规范（即 N 中的规则）的集合；Values(A) 表示 A 中涉及的所有价值的集合；TopRule(A) 表示 A 中所使用的最后一条规则。下文给出关于论辩系统中论证构造，以及关于上述集合的具体定义。

[2] 或者依惯例称之为普通前提集，参见 [14, 15]。

[3] 这在一些真实的论辩案例中有所体现，例如明代“大礼议”的案例，参见 [13] 中的分析。

定义 2.2 (论证). 设 $VAT=(\mathcal{L},\mathcal{K},V,\mathcal{R},n,val,\leqslant,\leqslant',P)$ 是一个考虑社会规范和价值的论辩理论，一个基于该论辩理论构造的论证 A 具有如下几种形式之一：

1. φ，如果 $\varphi\in\mathcal{K}$，并且 $\mathrm{Prem}(A)=\{\varphi\}$，$\mathrm{Conc}(A)=\varphi$，$\mathrm{Sub}(A)=\{\varphi\}$，$\mathrm{DefRules}(A)=\emptyset$，$\mathrm{Norms}(A)=\emptyset$，$\mathrm{Values}(A)=\emptyset$，$\mathrm{TopRule}(A)=undefined$；
2. $A_1,\ldots,A_n\longrightarrow\psi$，如果 $A_1,\ldots,A_n$ 都是论证，并且 $\mathcal{R}_s$ 中存在一条硬性规则 $\mathrm{Conc}(A_1),\ldots,\mathrm{Conc}(A_n)\longrightarrow\psi$，使得 $\mathrm{Prem}(A)=\mathrm{Prem}(A_1)\cup\ldots\cup\mathrm{Prem}(A_n)$，$\mathrm{Conc}(A)=\psi$，$\mathrm{Sub}(A)=\mathrm{Sub}(A_1)\cup\ldots\cup\mathrm{Sub}(A_n)\cup\{A\}$，$\mathrm{DefRules}(A)=\mathrm{DefRules}(A_1)\cup\ldots\cup\mathrm{DefRules}(A_n)$，$\mathrm{Norms}(A)=\mathrm{Norms}(A_1)\cup\ldots\cup\mathrm{Norms}(A_n)$，$\mathrm{Values}(A)=\mathrm{Values}(A_1)\cup\ldots\cup\mathrm{Values}(A_n)$，$\mathrm{TopRule}(A)=\mathrm{Conc}(A_1),\ldots,\mathrm{Conc}(A_n)\longrightarrow\psi$；
3. $A_1,\ldots,A_n\Longrightarrow\psi$，如果 $A_1,\ldots,A_n$ 都是论证，并且 $\mathrm{Norms}(A_1)\cup\ldots\cup\mathrm{Norms}(A_n)=\emptyset$，同时 $\mathcal{R}_d$ 中存在一条可废止规则 $\mathrm{Conc}(A_1),\ldots,\mathrm{Conc}(A_n)\Longrightarrow\psi$，使得 $\mathrm{Prem}(A)=\mathrm{Prem}(A_1)\cup\ldots\cup\mathrm{Prem}(A_n)$，$\mathrm{Conc}(A)=\psi$，$\mathrm{Sub}(A)=\mathrm{Sub}(A_1)\cup\ldots\cup\mathrm{Sub}(A_n)\cup\{A\}$，$\mathrm{DefRules}(A)=\mathrm{DefRules}(A_1)\cup\ldots\cup\mathrm{DefRules}(A_n)\cup\{\mathrm{Conc}(A_1),\ldots,\mathrm{Conc}(A_n)\Rightarrow\psi\}$，$\mathrm{Norms}(A)=\emptyset$，$\mathrm{Values}(A)=\emptyset$，$\mathrm{TopRule}(A)=\mathrm{Conc}(A_1),\ldots,\mathrm{Conc}(A_n)\Longrightarrow\psi$；
4. $A_1,\ldots,A_n\overset{v_u}{\Longrightarrow}\psi$，如果 $A_1,\ldots,A_n$ 都是论证，并且 N 中存在一条规范 $\mathrm{Conc}(A_1),\ldots,\mathrm{Conc}(A_n)\overset{v_u}{\Longrightarrow}\psi$，使得 $\mathrm{Prem}(A)=\mathrm{Prem}(A_1)\cup\ldots\cup\mathrm{Prem}(A_n)$，$\mathrm{Conc}(A)=\psi$，$\mathrm{Sub}(A)=\mathrm{Sub}(A_1)\cup\ldots\cup\mathrm{Sub}(A_n)\cup\{A\}$，$\mathrm{DefRules}(A)=\mathrm{DefRules}(A_1)\cup\ldots\cup\mathrm{DefRules}(A_n)$，$\mathrm{Norms}(A)=Norms(A_1)\cup\ldots\cup Norms(A_n)\cup\{\mathrm{Conc}(A_1),\ldots,\mathrm{Conc}(A_n)\overset{v_u}{\Longrightarrow}\psi\}$，$\mathrm{Values}(A)=\mathrm{Values}(A_1)\cup\ldots\cup\mathrm{Values}(A_n)\cup\{v_u\}$，$\mathrm{TopRule}(A)=\mathrm{Conc}(A_1),\ldots,\mathrm{Conc}(A_n)\overset{v_u}{\Longrightarrow}\psi$.

在上述论证中，构造过程中未使用规范的论证可以视为单纯用于实现认知推理的论证，而涉及规范和价值的论证可以视为用于实现实践推理的论证。包含不确定性元素的论证之间可能存在相互冲突，而仅由确定性元素（$\mathcal{K}_n$ 和 $\mathcal{R}_s$ 中的元素）构成的论证只可能攻击其他论证，但不能被其他论证攻击。

在得到论证集合之后，为了识别论证之间的冲突，我们给出如下关于论证间冲突（攻击关系）的定义。

定义 2.3 (论证间的攻击关系). 设 $VAT=(\mathcal{L},\mathcal{K},V,\mathcal{R},n,val,\leqslant,\leqslant',P)$ 是一个论辩理论，A、B 和 B' 是基于 VAT 所构造的论证。A 攻击 B，当且仅当 A 底切（undercut）、反驳（rebut）或者破坏（undermine）B 于 $B'\in\mathrm{Sub}(B)$，其中：

- A 底切 B 于 B'，当且仅当 $\mathrm{TopRule}(B')=r\in\mathcal{R}_d\cup N$，并且 $\mathrm{Conc}(A)=-n(r)$[4]；

[4] “$n(r)$”表示规则 r 是可应用的，“$-n(r)$”则表示使得该规则不可应用，从而在推理规则上对论证造成攻击。

- A 反驳 B 于 B'，当且仅当 B' 的形式为 $B"_1, \ldots, B"_n \Longrightarrow / \overset{v_u}{\Longrightarrow} \varphi$，并且 $\mathrm{Conc}(A) = -\varphi$；
- A 破坏 B 于 B'，当且仅当 B' 的形式为 φ，使得 $\varphi \in \mathrm{Prem}(B) \cap \mathcal{K}_p$，并且 $\mathrm{Conc}(A) = -\varphi$.

根据基于集合的优先级比较方法，例如文献 [9] 中所给出的“精英”（Elitist）和“民主”（Democratic）方法，由优先关系 $\leqslant$，$\leqslant'$ 和每位参与者所给出的 $\lesssim_i$，我们可以获得论证集合上的偏好关系。心理学、哲学、语用论辩等相关领域的研究成果中都曾指出，关于信念的知识应当优先于关于行动的决策得到辩护，从而保证后者能够建立在合理的信息和证据之上。（参见 [12, 19, 20]）因此，我们可以根据关于信念的优先关系 $\leqslant$ 和 $\leqslant'$，由合理的比较原则首先获得基于信念的论证偏好。在论辩系统定义中，常见的有“最弱链”和“最后链”的两种获取论证偏好的原则（[14]），前者考虑论证中所有的不确定性元素集合之间的优先关系，后者则考虑论证结构中最后端（最靠近结论）的不确定性元素集合之间的优先关系。其中，根据一些形式论辩理论研究者的分析，最弱链原则较适用于认知推理。（[15]）因此，在获得基于信念的论证偏好 $\prec$ 时，我们以最弱链原则为例，给出如下定义。

定义 2.4 (基于信念的论证偏好)**.** 设 A、B 是基于论辩理论 *VAT* 所构造的论证，根据集合之间的优先关系 $\lhd_s$（“s”表示某种提取优先关系的合理方法[5]），在最弱链原则下 $B \prec A$，当且仅当：

1. 如果 $\mathrm{DefRules}(B) = \emptyset$ 且 $\mathrm{DefRules}(A) = \emptyset$，那么满足 $\mathrm{Prem_p}(B) \lhd_s \mathrm{Prem_p}(A)$；
2. 如果 $\mathrm{Prem_p}(B) = \emptyset$ 且 $\mathrm{Prem_p}(A) = \emptyset$，那么满足 $\mathrm{DefRules}(B) \lhd_s \mathrm{DefRules}(A)$；
3. 否则，满足 $\mathrm{Prem_p}(B) \lhd_s \mathrm{Prem_p}(A)$ 且 $\mathrm{DefRules}(B) \lhd_s \mathrm{DefRules}(A)$.

$B \preceq A$，当且仅当 $B \prec A$，或者 $\mathrm{DefRules}(A) = \mathrm{DefRules}(B)$ 且 $\mathrm{Prem_p}(A) = \mathrm{Prem_p}(B)$.

相应的，最后链原则较适用于规范推理/实践推理，因此，可以根据最后链原则来获得基于社会规范和价值的论证偏好。由于本文假设了每个社会规范只关联一种主要的价值，每位参与者所给出的关于价值的排序 $\lesssim_i$ 可以对应于该参与者关于规范的排序。此外，对任意论证 A 而言，我们用 $\mathrm{LastNorms}(A)$ 表示包含规范的论证中的最后应用的规范的集合[6]，并将 *VAT* 中的“最后链”定义如下。

定义 2.5 (论证的“最后链”)**.** 设 A 是基于论辩理论 *VAT* 所构造的一个论证。如果 $\mathrm{Norms}(A) = \emptyset$，那么 $\mathrm{LastNorms}(A) = \emptyset$；如果 $A = A_1, \ldots, A_n \overset{v_u}{\Longrightarrow} \psi$，那么 $\mathrm{LastNorms}(A) = \{\mathrm{Conc}(A_1), \ldots, \mathrm{Conc}(A_n) \overset{v_u}{\Longrightarrow} \psi\}$；否则，$\mathrm{LastNorms}(A) = \mathrm{LastNorms}(A_1) \cup \ldots \cup \mathrm{LastNorms}(A_n)$.

[5] 例如上文所提到的“精英”或“民主”方法。（[9]）

[6] 根据定义 2.2，在一个论证的构造中使用规范后，普通的可废止规则不能应用在规范之后继续构造论证，因此本文中的最后链原则事实上是基于论证中最后应用的规范的集合进行比较。

为了与基于信念的论证偏好区分，我们用 $\prec_i^v$ 来表示基于社会规范和价值的论证偏好（其中 i 对应优先关系集合 P 当中的一个预序），并给出如下定义。

定义 2.6 (基于规范和价值的论证偏好)**.** 设 A、B 是基于论辩理论 *VAT* 所构造的论证，根据集合之间的优先关系 $\lhd_s$，对于每一个关于价值的优先排序 $\lesssim_i \in P$，在最后链原则下，$B \prec_i^v A$ 当且仅当：$B \prec A$，或者 $\mathrm{LastNorms}(B) \lhd_s \mathrm{LastNorms}(A)$ 且 $A \not\prec B$.

$B \preceq_i^v A$，当且仅当 $B \prec_i^v A$，或者 $\mathrm{LastNorms}(A) \neq \emptyset$ 且 $\mathrm{LastNorms}(B) = \mathrm{LastNorms}(A)$.

以文献 [13, 28] 中对明代“大礼议”案例的分析为例，以杨廷和为代表的“继嗣派”和以嘉靖帝等人为代表的论辩参与双方在“嘉靖帝是否继嗣于明孝宗”，以及“是否应当为嘉靖帝的亲生父亲上帝王尊号”这两个关键问题的争论中，对于对方所引用的典籍、传统文化和所推崇的价值本身并不存在相互攻击，可以说，他们在知识和信念问题上并不存在分歧，但分歧在于对“宗法”和“孝道”两种价值的优先排序。也就是说，如果“宗法”优先于“孝道”，那么相关的规范，如“根据传统典籍中对汉哀帝、宋英宗等继位过程的记载，嘉靖帝应该继嗣于明孝宗，而不应为亲生父亲上帝王尊号”（下文将这条规范简称为 n_1）应该优先于另一规范“根据《礼记》等传统典籍，嘉靖帝应该为亲生父亲上帝王尊号，而不应继嗣于明孝宗”（下文将这条规范简称为 n_2）；反之，如果“孝道”优先于“宗法”，则优先关系也相反。我们将 n_1 表示为“$t_1 \overset{v_z}{\Longrightarrow} s \wedge \neg g$”，$n_2$ 表示为“$t_2 \overset{v_x}{\Longrightarrow} g \wedge \neg s$”（其中“$t_1$”“$t_2$”分别用于表示不同的传统典籍，“$v_z$”“$v_x$”分别用于表示“宗法”和“孝道”两种价值，“$s$”“$g$”分别用于表示“嘉靖帝继嗣于明孝宗”和“为嘉靖帝的亲生父亲上帝王尊号”两种行动）。由此构造的两个论证 A_1 和 A_2 的结构分别如例 1 中所示。[7]

例 1 (包含规范的论证)**.**

$$A_1\colon t_1 \overset{v_z}{\Longrightarrow} s \wedge \neg g$$
$$A_2\colon t_2 \overset{v_x}{\Longrightarrow} g \wedge \neg s$$

根据定义 2.4 和定义 2.6，从信念上而言，论证 A_1 和 A_2 之间没有严格的偏好（$A \preceq B$ 并且 $B \preceq A$），但从规范和价值上而言，按“继嗣派”的观点，A_1 严格优先于 A_2，即 $A_1 \prec_1^v A_2$，而按嘉靖等人的观点，A_2 严格优先于 A_1，即 $A_2 \prec_2^v A_1$。

根据论证偏好，可以决定论证之间的攻击是否成功，成功的攻击可以称之为“击败”（defeat），定义如下。

[7] 本例只演示论证结构，未包含根据论证理论能够构造的所有论证和子论证。基于一个形式论辩系统的完整刻画过程可参考 [23, 28] 中的例示。

定义 2.7 (论证间的击败关系)**.** 设 A、B 和 B' 是基于一个形式论辩理论 *VAT* 所构造的论证，并且 $B' \in \mathrm{Sub}(B)$. 根据论证间的偏好 $\prec$ 以及 $\prec_i^v$（i 对应 $\lesssim_i \in P$），

1. A 基于信念击败 B，当且仅当 A 底切 B 于 B'，或者 A 反驳或破坏 B 于 B' 并且 $A \not\prec B'$；
2. A 基于价值击败 B，当且仅当：1）要么 A 基于信念严格地击败 B（即根据 $\prec$，A 击败 B 而 B 不能击败 A）；或者 2）A、B 基于信念互相击败，并且 $A \not\prec_i^v B'$.

我们用 D_i 表示击败关系的集合[8]，$(A, B) \in D_i$ 表示论证 A 击败论证 B。

根据上述定义，如果基于信念能够解决两个论证之间的相互冲突，即使得其中一个论证能够严格地击败另一个论证，那么他们之间的偏好就由 $\prec$ 决定；如果基于信念不能解决二者之间的相互冲突，那么他们之间的偏好进一步由 $\prec_i^v$ 决定。这也符合信念冲突优先于行动决策解决的预设。

根据论证集合和击败关系的集合，可以得到 Dung 式的抽象论辩框架（[10]），并由此进行论证状态评价，得到可同时被接受的论证集合。由于本文主要考虑的是在论辩参与者对于价值的优先排序不同的情况下，他们之间如何达成共识，参考 [10, 14]，下文基于可接受论证集的最基本要求，即“可相容”（admissible），给出如下关于共识的定义。

定义 2.8 (基于可接受论证集的共识)**.** 设 $\mathit{VAT} = (\mathcal{L}, \mathcal{K}, V, \mathcal{R}, n, val, \leqslant, \leqslant', P)$ 是一个考虑社会规范和价值的论辩理论，$F_i = \langle \mathcal{A}, D_i \rangle$ 是据此得到的一个抽象论辩框架，其中 $\mathcal{A}$ 是基于 *VAT* 所构造的所有论证的集合，D_i（i 与 $\lesssim_i \in P$ 对应）是 $\mathcal{A}$ 中的论证之间一组击败关系的集合，那么

- 一组论证的集合 $E \subseteq \mathcal{A}$ 是可相容的，当且仅当：1）$\nexists A, B \in E$ 使得 $(A, B) \in D_i$；2）$\forall B \in \mathcal{A}$，如果 $(B, A) \in D_i$，那么 $\exists X \in E$，使得 $(X, B) \in D_i$；$\mathcal{AD}_i$ 表示所有基于 F_i 得到的可相容论证集的集合；
- 对于任意基于 P 中的 $\lesssim_j$ 和 $\lesssim_k$ 得到的抽象论辩框架 F_j 和 F_k （$j \neq k$），如果 $\exists E$ 使得 $E \in \mathcal{AD}_j$ 且 $E \in \mathcal{AD}_k$，那么我们就说 $C = \{\mathrm{Conc}(A) | A \in E\}$ 是论辩参与者 j 和 k 之间的一个可相容共识.

除可相容集以外，在与论辩语义相关的研究中，研究者们给出了多种经典论辩语义。依据这些论辩语义，可以决定哪些论证“可辩护”（justified），以及哪些论证应被“拒斥”（rejected）。不同的语义可以适用于不同的推理情境与需求，有兴趣的读者可以参考 [3, 10] 等文献。

下面我们仍以“大礼议”之争为例进行考虑。在“大礼议”事件中，由于论争的双方在信念元素上没有分歧，因此即使无法在“嘉靖帝是否继嗣于明孝宗”，以

[8] 由于对于不同的参与者而言，由价值排序而获得的论证偏好可能不同，所得到的击败关系也不同，因此 D 也用下标作为区分，同样的，“i”对应 P 中的 $\lesssim_i$。

及“是否应当为嘉靖帝的亲生父亲上帝王尊号”这两个问题上达成共识，但双方都可以接受对方所参照的《皇明祖训》《礼记》《春秋公羊传》等历史典籍，以及其中所体现的传统文化观点，也就是说，论辩参与者这些社会文化背景上的认识是一致的。由此可以发现，在一些辩论中，有些共识（例如 $C = \{t_1, t_2\}$）虽然符合定义 2.8 中的要求，但并未回答论辩参与者所真正希望解决的争议性问题。因此，我们还应当讨论在何种条件下，人们能够达成“有意义”的共识。本文认为，一个有意义的共识，应当至少可以对参与者所关注的一些争议性问题提供回答。例如，根据例 1 中的论证，如果最终得到一个共识 C，使得 $\{g, \neg s\} \subseteq C$，那么对于“大礼议”事件中的“继嗣派”或者嘉靖帝一方在争论中所关心的争议性问题是有所回答的，即“为嘉靖帝的亲生父亲上帝王尊号”，同时“嘉靖帝不继嗣于明孝宗”。

设 I 表示一个论辩中参与者所关注的争议性问题的集合，我们用 $A \hookrightarrow^v E$ 表示 A 击败 E 中的至少一个论证，用 $E \hookrightarrow^v A$ 表示 E 中至少有一个论证击败 A，并用 $E \hookrightarrow^v E'$ 表示至少有一个 E 中的论证击败至少一个 E' 中的论证。此外，令 $E^{\hookrightarrow^v} = \{A \in \mathcal{A} \setminus E | E \hookrightarrow^v A\}$ 表示被 E 所击败的不属于 E 的论证的集合，并用 $\mathrm{Conc}(E)$ 表示 E 中所有论证的结论的集合，$Cl_{\neg}(\mathrm{Conc}(E))$ 表示集合 $\mathrm{Conc}(E)$ 在否定关系下的闭包。基于定义 2.8，论辩参与者可以得到有意义的共识的条件可以表示如下。

定理 2.1. *设 $\mathcal{F}$ 是一组基于论辩理论 VAT= $(\mathcal{L}, \mathcal{K}, V, \mathcal{R}, n, val, \leqslant, \leqslant', P)$ 得到的抽象论辩框架的集合，$\mathcal{A}$ 是基于 VAT 构建所有论证的集合，D_i 是基于 $\lesssim_i \in P$ 的 $\mathcal{A}$ 中的论证之间击败关系的集合，$E \subseteq \mathcal{A}$ 是一组论证的集合，$A \in \mathcal{A}$ 是一个论证。存在一个对所有参与者而言都有意义的共识集合 $C = \mathrm{Conc}(E)$，当且仅当所有基于 VAT 的抽象论辩框架 $F_i \in \mathcal{F}$ 都满足以下条件：*

1. $\mathrm{Conc}(E) \cap I \neq \emptyset$；
2. $E \not\hookrightarrow^v E$；
3. $\forall A \in \mathcal{A} \setminus E$，如果 $A \hookrightarrow^v E$，那么 $A \in E^{\hookrightarrow^v}$.

证明. ($\Rightarrow$) 1. 由于 $C = \mathrm{Conc}(E)$, 如果 C 是一个有意义的共识集，那么 C 中应该至少包含一个 I 中的元素（肯定的或否定的），即 $\mathrm{Conc}(E) \cap I \neq \emptyset$；2. 假设 $E \hookrightarrow^v E$，那么 $\exists A \in E$ 使得 $A \hookrightarrow^v E$，也即 $\exists B \in E$ 使得 $(A, B) \in D_i$，根据定义 2.8 中的第 1 点，这与 E 是可相容集的条件 1）相矛盾，因此 $E \not\hookrightarrow^v E$；3. 假设 $\exists A \in \mathcal{A} \setminus E$ 使得 $A \hookrightarrow^v E$ 且 $A \notin E^{\hookrightarrow^v}$，那么 $\exists B \in E$ 使得 $(A, B) \in D_i$ 且 $E \not\hookrightarrow^v B$，根据定义 2.8 中的第 1 点，这与 E 是可相容集的条件 2）相矛盾，因此条件 3 得证；

($\Leftarrow$) 1. 说明 C 中至少包含一个 I 的否定闭包中的元素，那么它对争议性问题有所回答，是一个有意义的共识；2. 根据定义 2.8，如果 $\nexists A \in E$ 使得 $A \hookrightarrow^v E$，那么 $\nexists A, B \in E$ 使得 $(A, B) \in D_i^v$，符合 E 是可相容集的第 1）点要求；3. 即

$\forall X \in E$，如果存在 $A \in \mathcal{A} \setminus E$ 使得 $(A, X) \in D_i$，那么 $\exists B \in E$ 使得 $(B, A) \in D_i$，根据定义 2.8 的第 1 点，可知 E 是一个可相容集。因此，$C = \mathrm{Conc}(E)$ 是一个有意义的可相容共识。 □

上述定理参考了基于抽象论辩框架的“强制”（enforcement）方面的研究（参见 [21]），可以视为相关方法与论辩的现实应用相关联的一种尝试。

3　相关工作

目前而言，在形式论辩研究中结合人文社科领域研究的成熟成果尚不多见。从结合形式论辩和价值偏好的角度而言，Bench-Capon 及其合作者所提出的基于价值的论辩框架（*VAF*）与本文的思路最为相关。（[2, 4]）*VAF* 主要从实践推理的角度考虑，假定每个论证与一个价值相关联，并由价值之间的优先关系决定论证之间的偏好。本文所给出的论辩理论和 *VAF* 都在形式论辩系统中结合了价值排序方面的考虑，但最大的不同之处在于 *VAF* 是主要基于抽象论辩框架的拓展，在论证的结构上没有实际的考量，因此在价值的关联上也不考虑论证结构和具体的关联方式。从与实际论证的接近程度而言，*VAF* 依然在一个比较靠近顶层的位置。而本文所给出的方法主要基于结构化论辩理论（[14]），将价值与规范相关联，依据价值和规范的优先排序逐步提升获得论证之间的偏好，同时将认知推理和实践推理分别处理，可以对社会文化背景下的论证进行更精细的建模，并对结论（论辩参与者之间所达成的共识）的形成给出更具体的解释。

Prakken 在 [16] 中的工作考虑认知推理和实践推理，提出了一种融合的论辩语义，并认为在特定语境下，对关于信念的推理应该采用比较怀疑的（sceptical）态度，而对关于行为的实践推理应该是采用比较轻信的（credulous）态度。Amgoud 等人在 [1] 中同样区分了上述两种类型的推理和论证，并给出了基于抽象论辩框架实现论证评价和行动决策的方法，以及几种决策原则。在论证与推理类型的考虑上，本文与这两项工作非常一致，同样将论证分为作用为实现认知推理的论证和作用为实现实践推理的论证，二者区分的主要标志在于构造中是否包含了社会规范和社会价值。除了以结构化论辩框架为基础进行拓展以外，本文还参考了广义论证理论中的思想，更关注论辩过程中如何导向共识的问题；同时，在共识的问题上，本文主要关注所得到的共识是否能够就论辩参与者所关心的问题给出有实际意义的回答。

与本文作者及合作者目前已发表的一些相关工作对比，文献 [22, 23] 等同样从广义论证理论中撷取了灵感和启发，主要从语境的动态性角度进行讨论，着重刻画了论辩过程中语境的变化。本文当前所给出的形式论辩理论则更侧重于讨论社会文化背景下与价值与规范相关的实践推理问题，细化了关于基于信念的论证

偏好和关于价值的论证偏好的获取方式，对包含认知推理和实践推理的论辩过程和多主体非单调推理给出一种更一般的框架。

4 总结与展望

本文主要结合人工智能领域的形式论辩研究和从社会文化领域出发的广义论证理论思想，构建了一个嵌入社会规范和价值、能够应对多主体价值偏好的形式论辩理论，对两种交叉学科领域的研究方法和成果的融合做了一个一般性的例示。基于该论辩理论，可以结合形式和非形式逻辑的研究思路，刻画在某种社会文化背景和具体语境下，持有不同价值观的论辩参与方如何达成一种集体可接受的共识的过程，实现考虑社会规范和价值的多主体推理决策。

由于本文关注的是同一社会文化背景下的论辩，因此我们所给出的论辩理论假设参与者在知识和信念方面已经具有一定的共识，只是需要从几种可能的选择中，根据对知识的优先排序，通过论证评价得到可辩护的论证和相应的结论。这与广义论证理论中对论证定义的基本思路相契合。（[26]）在此基础之上，论辩参与者可能由于目标和价值观上的分歧，在一些行动决策上仍然无法达成共识。这种情况可能由于新的目标和价值的加入而发生改变，最终使得论辩系统能够在参与者各自可接受的价值取向范围内导出一致的结论，也就是关于行动的共识。文献 [13] 中基于广义论证理论对历史事件“大礼议”之争的分析也正展示了这一过程。

通过本文的研究可以进一步发现，从形式论辩和广义论证研究的关联和差异上来说，广义论证理论强调自下而上的研究方法，而形式论辩系统也试图勾勒出这样一个论证构建和评价的过程。常见的形式论辩系统工作流程主要从底层具体的知识出发，进行形式化处理和表示，到顶层基于抽象化的论证完成计算评估，最后返回底层并输出结论。通过论证之间的攻击和击败关系，论辩系统可以决定哪些论证是基于当前所掌握的信息而言可辩护的，同时，那些被拒斥的论证结论不会出现在系统输出的结论集中。由此，形式论辩系统也可以在一定程度上体现推理中的动态性和非单调性特点，与广义论证理论所强调的论证系统可变异性和可修正性观点不谋而合。然而，当前大多数论辩系统仍以西方逻辑观和主流论证理论为基础，只给出了一种依据规则进行论证构建的简略框架，事实上，这一构建过程依然是单调不可撤销的。此外，相比广义论证的研究方法，其论证评价过程主要关注论证之间抽象的交互，依此决定论证状态，而或多或少地忽视了论证本身的生成和构造方式，难以反映文化特点如何影响和融入论辩系统的问题，以及论辩过程或语篇中论证策略的使用和转换，对于语境因素的表达刻画能力也比较有限。针对上述问题，广义论证的研究程序中特别强调了数据的田野采集，通过细致的研究步骤，获取特定文化群体的论证规则，并在这些规则的基础上建构该文化群体的论证活动。（[26]）这一研究程序可以为包含本土文化特点的社会规范、知识和信念的获取提供指导。最后，在推理结果的合理性审查方面，形式论辩系

统主要依据理性公设的给定，检查论辩系统输出的结果是否符合一致性等逻辑理性标准。([8])广义论证理论则通过经验方法，继续检验说话者所使用的规则是否符合社会规范，关心听话者的理解和论证的说服力，既关注论证的使用过程，也关注论证的使用效果，并持续以开放的观点关注论证的未来发展。这也为形式论辩系统在动态性、可修正性等方面功能的改进提供了启发性的思路。

在广义论证理论的启发下，基于现有的成果，以推进人工智能论辩系统的本土化应用为目标，我们还可以考虑如下几个方面的进一步研究。

1. 根据广义论证理论所强调的语境动态性和论证的局部合理性，论辩参与者可能通过修改语境的方式实现达成共识的目的。因此，未来的工作可以在当前研究的基础上，考虑如何明确地体现语境的动态性特点，使论辩系统能够更好地建模语境更新时论证状态的变化，并总结变化规律、降低计算复杂度。
2. 本文目前所给出的关于共识达成的定义，采用了一种基于可相容论证集和可接受结论集的基本判断聚合策略。在这个问题上，下一步工作首先可以根据广义论证理论，选择更适用于特定文化背景和情境下的论辩语义；其次，在偏好和判断聚合的策略选择上也值得进一步探究。
3. 实践推理通常是目标驱动的，因此，本文以能否对论辩过程中参与方所争论的关键问题给出回答为判断标准，参考关于论辩语义的研究成果，给出了对于能够达成“有意义”的共识时的论辩框架特点描述。目前所使用的判断标准可以视为一个最基本的要求，结合广义论证理论对于社会文化群体中的共识的特点总结，我们可以继续推进关于满足哪些特点的共识是一种“好的”共识的讨论。
4. 人工智能可解释问题是当前备受关注的研究热点，基于符号逻辑的形式论辩理论本身在可解释方面具有优势，而广义论证研究中所蕴含的社会科学研究方法可以在这方面提供更多有价值的指导，使系统设计能够以符合人类认知过程和决策的方式，提供更直观、更容易被用户理解的解释。进而，对社会文化和背景因素的强调，有助于人工智能论辩系统所提供的解释对文化差异和本土规范和语境保持敏感，确保解释的适当性。

形式论辩以符号逻辑为基础，脱胎于人工智能领域，是一个交叉学科研究主题。当前，论证挖掘、知识图谱等新兴技术的发展为形式论辩研究的发展提供了新的资源和空间，也提出了进一步的要求。形式论辩的跨学科属性使其在多个人工智能热门研究话题上皆有可为的空间。本文旨在说明广义论证理论不仅可以为非形式论辩、文化与逻辑史等研究领域提供指导，也可以对人工智能论辩系统的一些改良和本土化应用提供参考和启示。结合以广义论证理论为代表的、强调论辩的社会文化属性的研究，有望使论辩理论在时代背景下、扎根于本土文化，更好地发挥所长。

参考文献

[1] L. Amgoud and H. Prade, 2009, “Using arguments for making and explaining decisions”, *Artificial Intelligence*, **173(3):** 413–436.

[2] K. Atkinson and T. Bench-Capon, 2021, “Value-based argumentation”, *Journal of Applied Logics*, **8(6):** 1543–1588.

[3] P. Baroni, M. Caminada and M. Giacomin, 2011, “An introduction to argumentation semantics”, *The Knowledge Engineering Review*, **26(4):** 365–410.

[4] T. Bench-Capon, 2003, “Persuasion in practical argument using value-based argumentation frameworks”, *Journal of Logic and Computation*, **13(3):** 429–448.

[5] P. Besnard *et al.*, 2014, “Introduction to structured argumentation”, *Argument & Computation*, **5(1):** 1–4.

[6] P. Besnard and A. Hunter, 2014, “Constructing argument graphs with deductive arguments: A tutorial”, *Argument & Computation*, **5(1):** 5–30.

[7] M. Caminada, S. Modgil and N. Oren, 2014, “Preferences and unrestricted rebut”, in A. Utka *et al.* (eds.), *Frontiers in Artificial Intelligence and Applications, **Vol. 266**: Human Language Technologies—The Baltic Perspective*, pp. 209–220, Amsterdam: IOS Press.

[8] M. Caminada and L. Amgoud, 2007, “On the evaluation of argumentation formalisms”, *Artificial Intelligence*, **171(5):** 286–310.

[9] C. Cayrol, V. Royer and C. Saurel, 1992, “Management of preferences in assumption-based reasoning”, in B. Bouchon-Meunier, L. Valverde and R. R. Yager (eds.), *IPMU’92 - Advanced Methods in Artificial Intelligence*, pp. 13–22, Berlin, Heidelberg: Springer.

[10] P. M. Dung, 1995, “On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games”, *Artificial Intelligence*, **77(2):** 321–357.

[11] A. J. García and G. R. Simari, 2014, “Defeasible logic programming: DeLP-servers, contextual queries, and explanations for answers”, *Argument & Computation*, **5(1):** 63–88.

[12] G. Harman, 1986, *Change in View: Principles of Reasoning*, Cambridge, MA: MIT Press.

[13] S. Ju, W. Liu and Z. Chen, 2020, “L’argumentation sur la titulature impériale dans la dynastie ming au prisme de la «théorie généralisée de l’argumentation»”, *Argumentation et Analyse du Discours*, **25:** 1–23.

[14] S. Modgil and H. Prakken, 2013, “A general account of argumentation with preferences”, *Artificial Intelligence*, **195:** 361–397.

[15] S. Modgil and H. Prakken, 2014, “The ASPIC$^+$ framework for structured argumentation: A tutorial”, *Argument & Computation*, **5(1):** 31–62.

[16] H. Prakken, 2006, “Combining sceptical epistemic reasoning with credulous practical reasoning”, in P. E. Dunne and T. J. M. Bench-Capon (eds.), *Frontiers in Artificial Intelligence and Applications, **Vol. 144**: Computational Models of Argument, Proceedings of COMMA 2006*, pp. 311–322, Amsterdam: IOS Press.

[17] H. Prakken and M. De Winter, 2018, “Abstraction in argumentation: Necessary but dangerous”, in S. Modgil, K. Budzynska and J. Lawrence (eds.), *Frontiers in Artificial Intelligence and Applications, **Vol. 305**: Computational Models of Argument, Proceedings of COMMA 2018*, pp. 85–96, Amsterdam: IOS Press.

[18] F. Toni, 2014, “A tutorial on assumption-based argumentation”, *Argument & Computation*, **5(1):** 89–117.

[19] D. N. Walton, 1990, *Practical Reasoning: Goal-Driven, Knowledge-Based, Action-Guiding Argumentation*, Savage, MD: Rowman & Littlefield Publishers.

[20] D. N. Walton, 1996, *Argumentation Schemes for Presumptive Reasoning*, Mahwah, NJ: Lawrence Erlbaum Associates,

[21] K. Xu and B. Liao, 2021, “Characterizing argumentation frameworks with an extension”, *Studies in Logic*, **14(6):** 41–67.

[22] Z. Yu, 2020, “A context-based argumentation framework with values”, *Proceedings of the 20th Workshop on Computational Models of Natural Argument, CMNA 2020*, pp. 1–10, Perugia, Italy (and online), September 8th.

[23] Z. Yu and S. Ju, 2021, “Getting consensus through a context-based argumentation framework”, *Logics for New-Generation AI, Proceedings of the First International Workshop*, pp. 132–145, Hangzhou,China.

[24] 何杨、鞠实儿，“逻辑观与中国古代逻辑史研究的史料基础”，*哲学动态*，2019 年第 12 期，第 107–114 页。

[25] 鞠实儿，“论逻辑的文化相对性——从民族志和历史学的观点看”，*中国社会科学*，2010 年第 181 卷第 1 期，第 35-47+221–222 页。

[26] 鞠实儿，“广义论证的理论与方法”，*逻辑学研究*，2020 年第 1 期，第 1–27 页。

[27] 鞠实儿，“作为元方法的广义论证理论”，*第二届广义论证理论学术研讨会暨第四届中国古代论辩学术研讨会*，兰州，2023 年 7 月。

[28] 鞠实儿等，*论证挖掘与论证形式化*，2022 年，北京：科学出版社。

[29] 廖备水，“论新一代人工智能与逻辑学的交叉研究”，*中国社会科学*，2022 年第 3 期，第 37–54+204–205 页。

（责任编辑：袁之）

How to Incorporate Local Social Values into AI Argumentation Systems

—Combining Generalized Argumentation Theory with Value-Based Practical Reasoning

Zhe Yu

Abstract

Real-world argumentation takes place within specific social and cultural contexts, involving social norms and values. In the realm of practical reasoning and decision-making, participants' disagreements often stem not only from differences in beliefs and common knowledge but also from distinct priority orderings over social values. This article combines contextualized formal argumentation theory to explore how local social values can be incorporated into AI argumentation systems to foster consensus among participants with diverse value priorities. Through the analysis of argumentation cases, the proposed modeling approach can be extended to realize value-based practical reasoning, as well as investigate the integration of formal argumentation theory and the Generalized Argumentation Theory to evaluate arguments for practical reasoning with consideration of norms and values.

Zhe Yu Institute of Logic and Cognition, Sun Yat-sen University
Department of Philosophy, Sun Yat-sen University
yuzh28@mail.sysu.edu.cn