PHASE III - Existing and potential mitigations (*Version 1.0 - December 2021)*

Category	Type of Harm	Potential Harm / Misuse / Abuse	Contextual Example / Evidence	Existing and Potential Mitigations
consequential services	Opportunity loss	**Language discrimination** Limited language versioning on C2PA-enabled tools, despite their focus on low-cost and global accessibility, leads to more limited access for marginal markets.	C2PA-enabled tools are likely to leave out languages with marginal markets. A parallel example is that of the continued use in Myanmar of Zawgyi as the dominant typeface used to encode Burmese language characters rather than Unicode, the international text encoding standard, resulting in technical challenges for many companies that provide mobile apps and services.	**Specifications** Manifest localization specifications to be added beyond 1.0. No other limitations have been identified on language versioning of C2PA implementations. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should guide potential multilateral cooperation for the promotion of a diverse C2PA ecosystem, including language-inclusive implementations.
		Digital divide/technological discrimination (1) Individuals and communities using older devices or operating systems as creators/consumers or using access to the internet via Free Basics or equivalent "affordable access" approaches that limit the websites and services a customer can access.	For example, existing experiences with gated/limited access to particular websites and tools via Free Basics program for "affordable access" from mobile operators in emerging markets. See also example above on **Educational discrimination** and limited language versioning.	**Specifications** Specifications do not preclude C2PA implementations in older devices and operating systems. Specifications are open, global and opt in. The specifications use open standards for which there are existing libraries in various programming languages across a range of devices and operating systems/environments. To facilitate access for individuals or communities who do not, or cannot, have access to x.509 certificates, the specifications allow for self-signing certificates. **Accompanying documentation and guidance** Minimum viable implementations guidance to be developed as a fallback for older devices and operating systems. Guidance for implementers includes recommendations on the use of a private credential store (also known as the "address book"). **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should guide potential multilateral cooperation for the promotion of a diverse C2PA ecosystem and encourage the development of simple products to meet claim generation and validation requirements in diverse environments. An ongoing harm asessment should inform the continuous development of the specifications to address issues that limit C2PA implementations in older devices and operating systems.
		Digital divide/technological discrimination (2) Individuals and communities without abiilty to access or use tools for compliance with system usage are excluded.	Financial costs involved in signing up to use different C2PA-enabled tools and software may exclude marginalized individuals and communities who cannot afford the cost. For example, exclusion of content creators without compliant x.509 certificates. Lack of literacy and access to education about the tool may also limit usage among marginalized populations.	
		Journalistic Freedom and Independence An abuse of the C2PA system to enforce journalistic identity in laws in a jurisdiction or demand additional information on media posted on social media leads to a reduction of media diversity and suppression of speech. Misuse of provenance datastores to track content or enforce restrictive laws on freedom of expression and do so with lack of effective remedy and/or exploitation of provenance datastores to track content, and curtail freedom of expression (e.g. political speech). See overlap with **Journalistic Plurality and Diversity**	An escalation of laws addressing 'fake news', misinformation/disinformation and social media globally includes laws that enforce registered identity as a journalist on social media or provide governmental right-to-reply, which are being used to suppress dissent and reduce journalistic freedom.	**Specifications** Specifications are open, global and opt in. If they are used, the C2PA provides features that can be used to protect confidentiality of personal information while still establishing the provenance of an asset, including anonymous and pseudonymous signing, redaction as an authorized action, use of update manifests with redacted information, and the use of W3C credentials. No sensitive information is required in C2PA workflows. **Accompanying documentation and guidance** User experience guidance provides recommendations to prevent inadvertent disclosure of information. Guidance for implementers highlights trusts and privacy considerations, including on the use of datastores: We recommend that claim generators that add soft binding assertions to an asset's manifest do so as an opt-in addition and not make it mandatory. Guidance also recommends that content creators be informed of the trade offs involved in using provenance datastores that allow for asset link-up with soft bindings; that is, on the one hand, identifying manifests that have become 'decoupled' from their associated assets, while on the other hand, privacy risks that may result from a soft binding link-up to an earlier manifest with, for example, redacted information. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the C2PA to proactively engage and lobby for legislation that avoids misuses, and establish parallel compliance mechanisms to ensure implementations comply with C2PA's Guiding Principles.
		Loss of choice/network and filter bubble Risk of exacerbating epistemic injustice (whose accounts and knowledge are heard, validated and trusted) reflecting power dynamics of access and privilege among consumers and producers (including media) in terms of which information gets C2PA signals. See overlap with **Digital divide, Language discrimination, Differential pricing for goods and services, et. al.**	Existing dynamics of how both accreditation systems and epistemic trust focus on professional experience and exclude non-professional, community, non-accredited and historically marginalized communities.	**Specifications** This is an overarching concern that is addressed in more detail in various sections of this document (see: Digital divide, language discrimination et. al). Feedback sessions have included a diverse group of stakeholders to address potential epistemic injustices, but ongoing specifications development should reflect a continuous harms, misuse and abuse assessment. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the C2PA to proactively engage with communities, civil society and governments to promote a diverse C2PA ecosystem.

Denial of	**Economic loss**	**Devaluation of individual expertise** C2PA-based technology displaces skilled fact-checkers/journalists from news organizations and civil society watchdogs.	Existing precedents of displacement/shifts in employment of a range of skilled and unskilled professions by automation.	**Accompanying documentation and guidance** Implementations could result in a job shift (e.g. automation), but guidance includes human-in-the-loop for certain processes (e.g. soft binding link up). It is likely in the short-run that C2PA implementations could expand the capacity of fact checkers/journalists to engage with more content. It is recommended that any more permanent displacement/shift be held until a consolidated stage of C2PA implementatios. An ongoing harm asessment should inform the continuous development of the specifications to address devaluation of individual expertise.
		Differential pricing for goods and services Price discrimination as a result of participation in the marketplace for creative content or journalistic content disproportionately excludes marginalized communities and non-mainstream media who do not have access to relevant tools, or cannot consistently use tools because of privacy or other reasons. See overlap with **Journalistic Plurality and Diversity; Digital Divide, Forced association (Requiring participation in the use of technology or surveillance to take part in**	Existing precedents of increased pricing for content with particular characteristics perceived as valuable in marketplace	**Specifications** Specifications are open, global and opt in. Specifications cannot be used in pirated software, but they use open standards for which there are existing libraries in various programming languages across a range of devices and operating systems/environments, which should facilitate the development of implementations that reflect needs of users of a diverse ecosystem. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should guide the C2PA to cooperate multilaterally to promote a diverse ecosystem that includes free/libre, accessible implementations, specially for critical usages and for marginalized communities and individuals.
		Increased abuse of systems of creative ownerships C2PA-enabled attribution supports more extensive copyright trolling based on analysis of C2PA data.	Existing precedents of copyright trolling.	**Specifications** The standard assertions defined for use in a C2PA manifest include opportunities to add information about the attribution, rights and licenses of the associated asset, but the specifications have not been designed to reflect the needs of legal use-cases, and do not interact with existing mechanisms for copyright control. Ongoing specifications development should reflect the needs of impacted stakeholders in terms of access and derived legal usages. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should guide the C2PA to address potential abuse of systems of creative ownership bolstered by the use of C2PA manifests.
		Creative ownership impersonation C2PA assertions are used to make an apparent claim of ownership on another creator's work.	C2PA-backed assets may be used to claim ownership, e.g. a corporate NFT actor decides to create NFTs from CC licensed images.	**Specifications** The specifications do not legally establish ownership. The specifications allow for provenance information at capture to offer trust signals, but there are certain scenarios (i.e. legacy media, update manifests or security breach) where malicious or erroneous ownership claims could happen. The trust model highlights that the consumer determines whether or not to trust an asset based on their trust of the signer. Ongoing specifications development should reflect developing threats and harms related to claim ownership. The standard assertions defined for use in a C2PA manifest include opportunities to add information about the attribution, rights and licenses of the associated asset, but the specifications have not been designed to reflect the needs of legal use-cases, and do not interact with existing mechanisms for copyright control. **Accompanying documentation and guidance** User experience guidance emphasizes on this by offering four levels of information and is wary of potential misunderstanding, including taking C2PA stamps as evidence of truth. **Non-technical and multilateral harms response actions** Making malicious ownership claims counter C2PA's Guiding Principles and redress mechanisms should be considered by implementers. It is important to highlight that C2PA provenance information is not an indicator of truth.

	Dignity loss	**Public shaming, malinformation and targeted exposure and harassment** This may mean exposing people's private, sensitive, or socially inappropriate material (for example via doxxing based on C2PA-derived data, or using media created with C2PA data). See overlap with **Interference with Private Life** and **Never Forgotten**	Taking current or historical sensitive data from online platforms/services/devices and using C2PA to target particular groups such as women or women's right groups/LGBTQIA+ groups based on traffic. This can be both indiviual data and aggregate data. For example: activists where LGBTQIA+ is criminalized being deanonymized from individual and aggregate Grindr data to reveal use of the app.	**Specifications** The C2PA provides features that can be used to protect confidentiality of personal information while still establishing the provenance of an asset, including anonymous and pseudonymous signing, redaction as an authorized action, use of update manifests with redacted information, and the use of W3C credentials. No sensitive information is required in C2PA workflows. **Accompanying documentation and guidance** User experience and implementer guidance recommend allowing creators to opt in into a C2PA workflow. If the creator does opt in, the implementation should effectively allow for the user to retain control of the information recorded, with particular sensitivity to the creator's identity and process. In addition, consent to gather and share information should be given by the creator based on transparent processes (e.g. via a preview of the information gathered before generating a claim). Guidance for implementers highlights trusts and privacy considerations, including giving content creators the capacity to opt in into using provenance stores, to redact and delete manifests from datastores, and to determines whether their content may be queried or not. For soft binding, it is recommended that a human verify asset link-ups. The specifications also establish that soft binding should not replace hard binding. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the C2PA to proactively engage and lobby for legislation that avoids misuses, and establish parallel compliance mechanisms to ensure implementations comply with C2PA's Guiding Principles.
	Liberty loss, discrimination and due process	**Augmented Policing and Surveillance (1)** If data from C2PA and other sources were to be aggregated, it could be used to target and discriminate against individuals and groups. This could occur for example with police body cams if they are equipped with facial recognition technologies that reinforce racial and other biases. C2PA data is used to amplify surveillance mechanisms and to infer suspicious behavior and/or criminal intent based on historical records.	Harms around facial recognition are now well documented in a law enforcement context. Data from C2PA could also be incorporated into invasive online identification schemes.	**Specifications** The C2PA provides features that can be used to protect confidentiality of personal information while still establishing the provenance of an asset, including anonymous and pseudonymous signing, redaction as an authorized action, use of update manifests with redacted information, and the use of W3C credentials. No sensitive information is required in C2PA workflows. C2PA 1.0 specs focused on embedded assertions rather than cloud-hosted assertions wherever possible - and structured to avoid doing this during normal playback requests. **Accompanying documentation and guidance** User experience and implementer guidance recommend allowing creators to opt in into a C2PA workflow. If the creator does opt in, the implementation should effectively allow for the user to retain control of the information recorded, with particular sensitivity to the creator's identity and process. In addition, consent to gather and share information should be given by the creator based on transparent processes (e.g. via a preview of the information gathered before generating a claim). Guidance for implementers highlights trusts and privacy considerations, including giving content creators the capacity to opt in into using provenance stores, to redact and delete manifests from datastores, and to determines whether their content may be queried or not. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the C2PA to proactively engage and lobby for legislation that avoids misuses, and establish paraleel compliance mechanisms for implementations that counter C2PA's Guiding Principles.
		Augmented Policing and Surveillance (2) Biometric identification approaches incorporate C2PA-spec'd devices for capture and enhanced protection for biometric scans for identification, resulting in additional data for identification of individuals, with potential privacy-compromising and impacts on both obligatory usage for marginalized-community as well as exclusion consequences.	Biometric and digital identity systems deployed without public accountability	
		Augmented Policing and Surveillance (3) Privacy loss for consumers of media via tracking of access to cloud-based assertion data (e.g. via tracking pixels)	Eg. tracking pixels of access to cloud-based assertion data.	
		Augmented Policing and Surveillance (4) Use of broader availability of provenance data to do broad search for content, e.g. geofenced location data search	Significant concern since journalistic/media usage of a C2PA-enabled search is an identified use case	
		Loss of Effective Remedy (1) Usage of C2PA signals in automated systems lack capacity to adequately interpret a complex set of complementary signals. Consumers lack transparency and right of appeal to potential algorithmic bias or harms from interpretation of C2PA signals.	Loss of effective remedy in existing systems of algorithmic recommendation or downranking, or algorithmic content removal. In C2PA, decoupled, hard bound manifests are automatically matched to its content. Although unlikely, hash collitions cannot be ruled out. C.f. literature on accuracy of automated systems in complex contextual situations - e.g. misinformation ("Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis", Center for Democracy and Technology") and on opacity in authenticity infrastructure	**Specifications** Within the specifications, loss of effective remedy could result from erroneous soft binding link-ups or from correct link-ups to previous, more data-rich manifests that include sensitive information. To address this, the specifications establish that 1. soft bindings cannot replace hard bindings, and 2. soft bindings are not required. **Accompanying documentation and guidance** For soft binding, it is recommended that a human verify asset link-ups. The specifications also establish that soft binding should not replace hard binding. Platform algorithmic recommendations and downranking should consider accessibility to C2PA-enabled tools, as well as potential harms deriving from policies that mandate its use, including considerations around certificate authorities and anonymous or pseudonymous signing of certificates. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the C2PA to proactively engage and lobby for legislation that avoids misuses, and establish parallel compliance mechanisms to ensure implementations comply with C2PA's Guiding Principles.
		Loss of Effective Remedy (2) In case of an inaccurate or misleading C2PA result, individuals will not have the ability to contest a technical decision that may have repercussions on a personal or community level	Existing forensic explainability challenges of media forensics - c.f. the controversy even over whether/how a World Press Photo prize-winning photo was manipulated.	

Infringement on human rights	**Privacy loss**	**Interference with private life (1)** Inadvertent disclosure of information (from unintended inclusion of assertions, or disclosure of assertions in an unintended way) or aggregate information from assertion combined with other data. See overlap with **Never Forgotten** and **Public Shaming**	Taking sensitive data from online platforms/services/devices and using C2PA to target particular groups such as women or women's right groups/LGBTQIA+ groups based on traffic. For example: activists where LGBTQIA+ is criminalized being deanonymized from Grindr data to reveal use of the app.	**Specifications** The C2PA provides features that can be used to protect confidentiality of personal information while still establishing the provenance of an asset, including anonymous and pseudonymous signing, redaction as an authorized action, use of update manifests with redacted information, and the use of W3C credentials. No sensitive information is required in C2PA workflows. C2PA 1.0 specs focused on embedded assertions rather than cloud-hosted assertions wherever possible - and structured to avoid doing this during normal playback requests. **Accompanying documentation and guidance** User experience and implementer guidance recommend allowing creators to opt in into a C2PA workflow. If the creator does opt in, the implementation should effectively allow for the user to retain control of the information recorded, with particular sensitivity to the creator's identity and process. In addition, consent to gather and share information should be given by the creator based on transparent processes (e.g. via a preview of the information gathered before generating a claim). Guidance for implementers highlights trusts and privacy considerations, including giving content creators the capacity to opt in into using provenance stores, to redact and delete manifests from datastores, and to determines whether their content may be queried or not. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the C2PA to proactively engage and lobby for legislation that avoids misuses, and establish parallel compliance mechanisms to ensure implementations comply with C2PA's Guiding Principles.
		Interference with private life (2) Misuse of C2PA soft-binding for broadened scope of data-rich searchable hash stores outside of traditional 'violating' hash databases including use in emergent client-side data scanning in encrypted messaging.	Recent experience with Apple's plans to scan client-side photos for Child Sexual Abuse Material (CSAM). Although it was meant to ensure child safety, it had broad and potentially harmful implications. See International Coalition Calls on Apple to Abandon Plan to Build Surveillance Capabilities. Existing client-side scanning in PRC	
		Reduction in options for anonymity and pseudonymity Inadvertent disclosure of information (from unintended inclusion of assertions, or disclosure of assertions in an unintended way) or aggregate information from assertion combined with other data.	Human rights activist inadvertently includes location in media assertion and is subsequently targeted (c.f. existing precedents of inadvertent release of metadata) Aggregate mobile data provides the ability to deanonymize individuals, for example through phone reversal lookup APIs, or through public records search or breach in the case of countries where biometric data is required for telecommunication services (c.f. Mexico reform to the federal telecommunications law).	**Specifications** Specifications are open, global and opt in. If they are used, the C2PA provides features that can be used to protect confidentiality of personal information while still establishing the provenance of an asset, including anonymous and pseudonymous signing, redaction as an authorized action, use of update manifests with redacted information, and the use of W3C credentials. No sensitive information is required in C2PA workflows. **Accompanying documentation and guidance** User experience and implementer guidance recommend allowing creators to opt in into a C2PA workflow. If the creator does opt in, the implementation should effectively allow for the user to retain control of the information recorded, with particular sensitivity to the creator's identity and process. In addition, consent to gather and share information should be given by the creator based on transparent processes (e.g. via a preview of the information gathered before generating a claim).
		Never forgotten Digital files or records may never be deleted. Depending on a given C2PA-enabled system's functionality for redaction of soft or hard binding provenance datastores, and/or what usage of irrevocable ledgers, risk of digital files that contain misinformation OR that contain privacy and dignity-compromising information continues to circulate. Storage of manifests may not allow for manifest redaction, deletion or selective disclosure. A content creator may want to delete private/sensitive manifest from content, but decoupled manifests in provencnace data stores may reattach manifest to content (c.f. analogous to Google/Facebook image storage). Further questions around the capacity to redact information in blockchain-based C2PA systems remain.	For example: human rights defenders, journalists and others in Afghanistan removing content showing their face and personal information from Internet/social media	**Specifications** The C2PA provides features that can be used to protect confidentiality of personal information while still establishing the provenance of an assetl including anonymous and pseudonymous signing, redaction as an authorized action, use of update manifests with redacted information, and the use of W3C credentials. No sensitive information is required in C2PA workflows. C2PA 1.0 specs focused on embedded assertions rather than cloud-hosted assertions wherever possible - and structured to avoid doing this during normal playback requests. **Accompanying documentation and guidance** User experience and implementer guidance recommend allowing creators to opt in into a C2PA workflow. If the creator does opt in, the implementation should effectively allow for the user to retain control of the information recorded, with particular sensitivity to the creator's identity and process. In addition, consent to gather and share information should be given by the creator based on transparent processes (e.g. via a preview of the information gathered before generating a claim). Guidance for implementers highlights trusts and privacy considerations, including giving content creators the capacity to opt in into using provenance stores, to redact and delete manifests from datastores, and to determines whether their content may be queried or not. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the C2PA to proactively engage and lobby for legislation that avoids misuses, and establish paraleel compliance mechanisms for implementations that counter C2PA's Guiding Principles.

	Constraints on Freedom of Expression	**Inability to freely and fully develop personality and creative practice** Workplace requirements to use tools for production in journalistic/creative contexts may have implications for personal privacy and personal artistic practice by forcing disclosure of techniques	Media and artistic/creative producers are concerned about disclosure of creative techniques	**Specifications** C2PA specifications do not require disclosure of creative techniques (actions). However, once included in a manifest, these actions may not be redacted without rendering the manifest invalid, or generating a new claim. **Accompanying documentation and guidance** User experience and implementer guidance recommend allowing creators to opt in into a C2PA workflow. If the creator does opt in, the implementation should effectively allow for the user to retain control of the information recorded, with particular sensitivity to the creator's identity and process. In addition, consent to gather and share information should be given by the creator based on transparent processes (e.g. via a preview of the information gathered before generating a claim). **Non-technical and multilateral harms response actions** A continuous harms, misuse and abuse assessment should consider feedback from creators community to understand how the specifications may hinder their creative practices and raise privacy risks in order to adjust specifications accordingly.
		Enforcement of extralegal or restrictive laws on freedom of expression Misuse of provenance datastores to track content or enforce restrictive laws on freedom of expression and do so with lack of effective remedy and/or exploitation of provenance datastores to track content, and curtail freedom of expression (e.g. political speech)	Political dissidents being tracked through C2PA datastores, or 'bad actors' demanding datastores to release sensitive information.	**Specifications** In addition to the features provided to protect the confidentiality of personal information, the specifications do not require storing manifests remotely, and if they are, it does not require that manifests be available for public querry, in particular through soft binding, which raises some privacy concerns. **Accompanying documentation and guidance** User experience and implementer guidance recommend allowing creators to opt in into a C2PA workflow. If the creator does opt in, the implementation should effectively allow for the user to retain control of the information recorded, with particular sensitivity to the creator's identity and process. In addition, consent to gather and share information should be given by the creator based on transparent processes (e.g. via a preview of the information gathered before generating a claim). Harms modelling also highlights the importance of considering local legislation and context for implemenations and datastores that may be abused to track content and curtail freedom of expression. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should continuously inform the specifications development process to mitigate misuses and abuses of the C2PA to enforce extralegal or restrictive laws on freedom of expression. In addition, the C2PA should drive efforts to establish parallel compliance mechanisms *t*o ensure implementations comply with C2PA's Guiding Principles.
		Forced association (Requiring participation in the use of technology or surveillance to take part in society) De facto inclusion and participation obligation in marketplaces for creative content or journalistic content or for better algorithmic ranking on social media sites which disproportionately excludes global populations, marginalized communities and non-mainstream media who do not have access to relevant tools, or cannot consistently use tools because of privacy or other reasons.	For example, algorithmic ranking: content creators forced to game algorithms with particular keywords, metadata to achieve visibility/to be ranked higher in a feed.	**Specifications** Specifications are open, global and opt in. Continuous specifications development should reflect needs of users of a diverse ecosystem. **Accompanying documentation and guidance** Although some implementations may require its users to generate C2PA manifests, it is not expected that services or products with a broader use-base (e.g. certain social media platforms) will require C2PA-embedded assets as part of their workflow. Future guidance should include scenarios for a wide adoption stage, including use in social media platforms. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should guide the C2PA to cooperate multilaterally to promote a diverse ecosystem that includes free/libre, accessible implementations, specially for critical usages and for marginalized communities and individuals. It should also offer guidance to social media platforms or other implementers whose tools may have broader societal or industry-wide impacts.

	Freedom of Association, Assembly and Movement	**Loss of freedom of movement or assembly to navigate the physical or virtual world with desired anonymity** C2PA-enabled systems that utilize a real-name identity or other real-world profile provide a mechanism to connect movement in space to an individual via C2PA metadata	Inadequate UX or implementation creates simplistic signals of trust that obscure real-life dynamics faced by individuals who mix some elements/moments of public visibility with pseudonymity and anonymity in other circumstances.	**Specifications** The C2PA provides features that can be used to protect confidentiality of personal information while still establishing the provenance of an assetl including anonymous and pseudonymous signing, redaction as an authorized action, use of update manifests with redacted information, and the use of W3C credentials. No sensitive information is required in C2PA workflows. C2PA 1.0 specs focused on embedded assertions rather than cloud-hosted assertions wherever possible - and structured to avoid doing this during normal playback requests. **Accompanying documentation and guidance** User experience and implementer guidance recommend allowing creators to opt in into a C2PA workflow. If the creator does opt in, the implementation should effectively allow for the user to retain control of the information recorded, with particular sensitivity to the creator's identity and process. In addition, consent to gather and share information should be given by the creator based on transparent processes (e.g. via a preview of the information gathered before generating a claim). Harms modelling also highlights the importance of considering local legislation and context for implemenations and datastores that may be abused to track content and curtail freedom of expression. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the C2PA to proactively engage and lobby for legislation that avoids misuses, and establish parallel compliance mechanisms to ensure implementations comply with C2PA's Guiding Principles.
	Environmental impact	**High energy consumption** Extensive use of blockchain or types of distributed ledger technology with C2PA-enabled systems contributes to exploitation of natural resources.	Blockchain-enabled C2PA systems would be part of a broader high-energy consumption ecosystem. The assessment however reflects the assumption that these systems would primarily operate with proof-of-stake models.	**Specifications** Specifications do not require or preclude the use of blockchain-enabled C2PA systems. **Accompanying documentation and guidance** To address privacy concerns, it is recommended that C2PA manifests should not be stored on DLTs. DLTs could be used to underwrite the integrity of a datastore containing C2PA manifests (for example a cloud database). Proof of Stake DLTs could be used so that energy use is relatively low/in-line with typical cloud computing energy usage.
res	**Manipulation**	**Misinformation (1)** C2PA-enabled systems can be used to generate misinformation (for example by generation of deliberately misleading manifests) and imply that it is trusted.	*[From security considerations]* An attacker misuses a legitimate claim generator (e.g. C2PA-enabled photo editor) to add misleading provenance to a C2PA-enabled media asset. See Journalistic plurality and diversity	**Specifications** The specifications establish a trust model; C2PA's commitment is to provide signals of trust, and not to arbitrate or confirm the integrity of assets or to determine truth. A threats assessment has been carried out to strive towards the integrity of the system by providing security features and considerations to prevent and mitigate threats and harms. For a detailed threats analysis, see security considerations documentation. The development of the specifications should reflect ongoing threats and harms assessments that reflect mis/dis/malinformation concerns and impact. **Accompanying documentation and guidance** Mis/dis/malinformation may result from the misuse or abuse of signing systems. To address this, guidance for implementers includes recommendations on how to protect claim signing keys and how to verify the suitability of signing credentials, including revocation and time stamp guidance. The C2PA does not mandate the use of any specific list of certificates or CAs that can be used to verify the trustworthiness of the signer of a manifest, but it recognizes that harms could arise from this if the implementations are not careful to address some of the issues listed in this document (and others that may arise). User experience guidance aims to define best practices for presenting C2PA provenance to consumers, this includes considerations on potential scenarios for mis/dis/malinformation. It also establishes that, rather than attempt to determine the veracity of an asset, it should enable users to make their own judgement by presenting the most salient and/or comprehensive provenance information. For more details on this, see User experience guidance. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should inform the development of the specifications to address a changing landscape and other unforeseen threats and harms. Additionally, the C2PA recognizes that bad actors may be willing and capable of creating their own tools that forego the specifications and its accompanying guidance. To address this, the C2PA should proactively engage and lobby for legislation that avoids misuses, to help establish parallel compliance mechanisms, and to cooperate for a diverse C2PA ecosystem.
		Misinformation (2) C2PA-enabled tools can be used to support misinformation, and in certain circumstances make it hard to revoke or retract this misinformation, leaving a continued assumption of additional trust.	An erroneous C2PA manifest is used to "validate" an image that is widely shared. C2PA manifest is not revoked or retracted, so the image is continuouslly shared and trusted.	
		Misinformation (3) Disguising fake information as legitimate or credible information by deliberate mis-attribution and assignation of C2PA provenance to existing content (without C2PA data) and legacy media, and addition of relevant soft and hard bindings to provenance datastores where look-up provides deceptive results on first visual glance.	For example, if using thumbnails or other low quality images to do a soft binding look up of an asset throws back a wrong match. This information could then be used to misinform. C2PA validator fosters a loss of remedy in cases like these.	
		Misinformation (4) C2PA-enabled ecosystem creates an 'implied falsehood' around media that does not contain C2PA assertions/manifests, resulting in discrediting of legitimate content sources.	Videos or images from sources that cannot or prefer to not used C2PA-enabled devices are discredited or undermined.	
		Misinformation (5) Mislabeling trustworthy information as misinformation: C2PA-enabled tools and derived signals AND/OR soft-binding hashes can be used inappropriately in automated systems for detecting, classifying, organizing, managing and presenting misinformation.	C.f. analysis on automation of visual misinformation detection	

Erosion of Social and Democratic Structu	**Over-reliance on systems**	**Overconfidence in technical signals** Over-confidence in the technical signals as an indicator of truth or confirmation of trust, rather than a set of signals related to provenance and authenticity/edits. Use of automated look-up systems progressively reduces human-in-the-loop, leading to exacerbated problems around contextualization of information or augmentation of problems (for example, mislabeling of misinformation based on contextual misunderstanding, or from malicious uses articulated above). These problems could occur at the front-end providing deceptive UX assumptions to soft-binding look-up or at back-end with over-automation of usage of soft-binding signals. See overlap with **Loss of remedy** and automation	C.f. literature on overconfidence in simple technical signals, particularly in misinformation systems. In C2PA specifications, an open question remains on the issue of automatically linking digital assets to manifests in provenance databases through soft binding matches.	**Specifications** The specifications establish a trust model; C2PA's commitment is to provide signals of trust, and not to arbitrate or confirm the integrity of assets or to determine truth. Look-up systems using soft bindings for decoupled manifests could produce errors or be subject to attacks. To address this, the specifications establish that 1. soft bindings cannot replace hard bindings, and 2. soft bindings are not required. **Accompanying documentation and guidance** It is recommended that matches made using a soft binding must be interactively verified via human-in-the-loop. It is also recommended that claim generators that add soft binding assertions to an asset's manifest do so as an opt-in addition and not make it mandatory. To mitigate risks to user privacy, it is recommended thath content creators be informed of the trade offs involved in using provenance datastores that allow for asset link-up with soft bindings; that is, on the one hand, identifying manifests that have become 'decoupled' from their associated assets, while on the other hand, privacy risks that may result from a soft binding link-up to an earlier manifest with, for example, redacted information. **Non-technical and multilateral harms response actions** The C2PA should drive efforts to highlight that its commitment is to provide signals of trust, and not to arbitrate or confirm the integrity of assets or to determine truth. The ongoing harms, misuse and abuse assessment should inform of further potential mitigations as implementations are rolled out.
	Social detriment	**Amplification of power inequality** Requiring participation in the use of technology to take part in society.	De facto inclusion and participation obligation in marketplaces for creative content or journalistic content or for better algorithmic ranking on social media sites which disproportionately excludes global populations, marginalized communities and non-mainstream media who do not have access to relevant tools, or cannot consistently use tools because of privacy or other reasons.	**Specifications** Specifications are open, global and opt in. Some considerations at the specifications level: claim generators are not expected to work with pirated software; specifications can be used entirely offline; specifications use open standards for which there are existing libraries in various programming languages across a range of devices and operating systems/environments. Continuous specifications development should reflect needs of users of a diverse ecosystem. **Accompanying documentation and guidance** Amplification of power inequalities may result from lack of access to C2PA implementations or from legal or de facto requirements to use them. Guidance is offered to create inclusive implementations. User experience guidance is also offered to define best practices for presenting C2PA provenance to consumers, including to avoid disinfranchising voices that may -or choose to- not use C2PA implementations. **Non-technical and multilateral harms response actions** The harms, misuse and abuse assessment should guide the C2PA to inform key actors (e.g. major social media) of the risks of establishing required use of the C2PA, especially in an initial adoption stage, where implementations are not accessible to all. Additionally, the C2PA should drive efforts to resource and promote a diverse C2PA ecosystem that addresses the need of a broad range of individuals and communities.
		Journalistic plurality and diversity "A divergence of usage between media able to afford/adapt to/use C2PA-enabled tools and workflows, and a broader range of smaller media and individual citizen journalists leads to a de facto two tier trust system in public perception." See overlap with **Journalistic Freedom and Independence**	Smaller or community news publishers are unable to provide C2PA-backed content, and so their content is undermined by audiences, platforms, governments, influential individuals (eg. Liar's dividend).	**Specifications** To facilitate implementations for civic, community and independent media, it should be noted that the specifications are open, global and opt in, and that it uses open standards for which there are existing libraries in various programming languages across a range of devices and operating systems/environments. To facilitate the use of C2PA implementations for civic/community/independent media, the specifications allow for self-signing certificates for those that do not, or cannot, have access to x.509 certificates.
	Emotional or psychological distress; Physical harm	**Misattribution and Malinformation (2)** Misuse of C2PA-enabled media to implicate an individual or group in inciting violence/criminality, or otherwise negatively or positively impact the reputation of a group or individual or media entity. Including deliberate mis-attribution and assignation of C2PA provenance to existing content (without C2PA data) and legacy media, and addition of relevant soft and hard bindings to provenance datastores where look-up provides deceptive results on first visual glance (e.g. thumbnail approach) See overlap with **Never Forgotten**	Existing problems of digital wildfire (rapidly-shared online content) frequently feature existing shallowfaked, miscontextualized content claimed to be from one place when actually from another. Patterns of manipulating media to misattribute are commonplace and should be assumed as an attack vector for C2PA-enabled systems.	**Specifications** The specifications establish a trust model; C2PA's commitment is to provide signals of trust, and not to arbitrate or confirm the integrity of assets or to determine truth. Look-up systems using soft bindings for decoupled manifests could produce errors or be subject to attacks. To address this, the specifications establish that 1. soft bindings cannot replace hard bindings, and 2. soft bindings are not required. **Accompanying documentation and guidance** It is recommended that matches made using a soft binding must be interactively verified via human-in-the-loop. It is also recommended that claim generators that add soft binding assertions to an asset's manifest do so as an opt-in addition and not make it mandatory. To mitigate risks to user privacy, it is recommended thath content creators be informed of the trade offs involved in using provenance datastores that allow for asset link-up with soft bindings; that is, on the one hand, identifying manifests that have become 'decoupled' from their associated assets, while on the other hand, privacy risks that may result from a soft binding link-up to an earlier manifest with, for example, redacted information.