Digital policy analysis: A methodological toolkit

Ellen J. Helsper, Shivani Rao and Marisa Lyons Longworth

Addendum by Anran Wang

December 2025

Digital Futures For Children

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

5RIGHTS FOUNDATION

Table of contents

Acknowledgements

We would like to express our gratitude for the support of the Digital Futures for Children centre (DFC) and 5Rights, whose funding made this research and the associated toolkit possible. We would also like to thank Anran Wang for his work on the Addendum related to cosine similarity mapping.

The contributions the research assistants made to developing and finetuning the codebooks and analytical frameworks were invaluable. Thank you, Ghea Adasta-Fauzan, Omer Gibreel, Laura Betancourt-Basallo, Thomas Cessou, Helen Ho, Malika Kaloo, Diego Martínez-Castro, Andri Ridhiani Putri, Zayna Resley, Sramana Sabnam, Thaís Sanches Cardoso, Ruofan Shi, Tham Truong, Tanishqa Vaish, Remi Waxman, Cong Xu and Hao-Yun Yu.

What does this toolkit do?

This toolkit provides researchers with a step-by-step guide detailing a systematic methodology for evaluating if and how digital policies represent certain groups and their rights.

The methodology underpinning the toolkit facilitates an in-depth understanding of how policies consider diversity and inequalities, identifying opportunities for improvement in policies that are shaping not just the present, but also the future.

This toolkit can be adapted and applied across different political, legal and cultural contexts and policy domains. It can be used to analyse policies at the global, regional, national, local and institutional level, to uncover commonalities and divergences between and within policies at these levels and between and within countries.

This toolkit is borne out of research conducted for the Digital Futures for Children centre (DFC).[1] The DFC project analysed the representations of children, their rights and inequalities between them in over 300 digital policies across 35 intergovernmental, regional and national governing bodies.

For a detailed description of the findings, please read the accompanying reports: *Left out and misunderstood: Children in digital policies. A global review*[2] and *Left out and misunderstood: Children in global, regional and national digital policies.*[3]

[1] Hereafter referred to as the 'DFC project'.
[2] Helsper et al. (2025a).
[3] Helsper et al. (2025b).

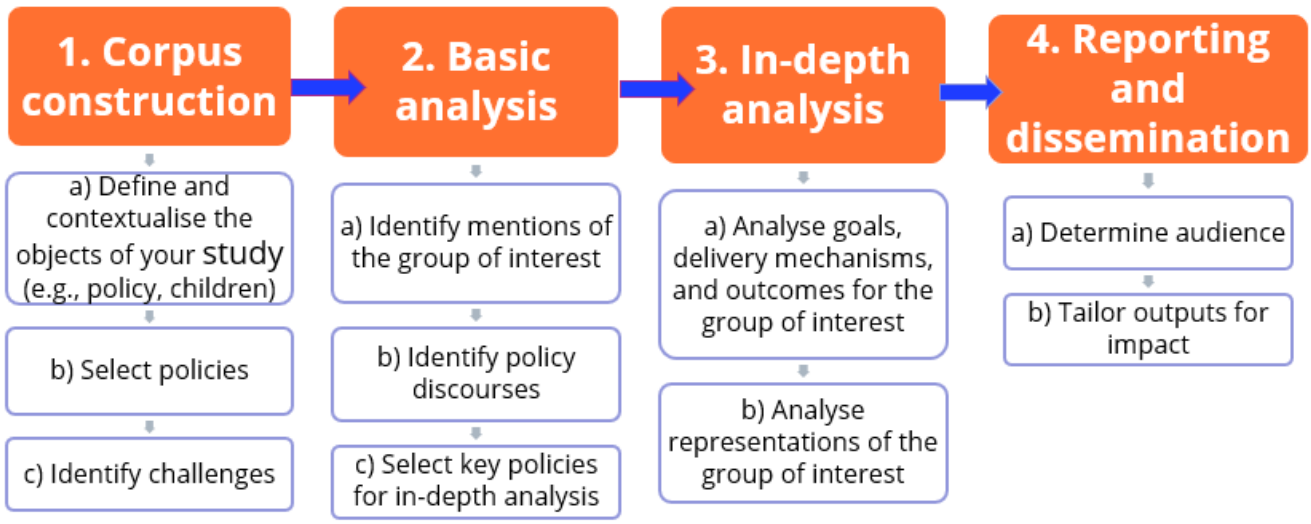
Who is this toolkit for?

- **Researchers seeking systematic methods** for the analysis and evaluation of digital policies, for example those interested in child and human rights, digital inequality and digital governance.
- **International and intergovernmental organisations** conducting comparative policy assessments across countries, supporting member states and development agencies in policy development, and holding them accountable for implementing treaties or goals.
- **Government departments, researchers and audit offices** evaluating the effectiveness of digital policies and developing new digital policies based on this.
- **Independent oversight bodies, courts, lawyers and human rights bodies** assessing policy compliance, holding governments to account.
- **Human and digital rights advocates**, civic society and local campaigners, journalists and others examining how digital policies affect specific groups in the regions they cover, and building evidence-based campaigns for policy and social reform.
- **Child rights organisations**, providers of children's services trying to understand and evaluate the impact of global, national and local digital policies on children's futures and current rights.
- **Industry associations** developing self-regulatory frameworks, **corporate social responsibility** and **equity and diversity** teams evaluating digital policies' impact on specific groups, **tech policy consultants** advising on human and child rights compliance, **EdTech specialists** ensuring child-centred approaches, etc.

Step-by-step guide to digital policy analysis

This toolkit is organised around the four steps we suggest researchers, evaluators and designers take when conducting digital policy analysis (see Figure 1).

Figure 1: Four-step policy selection and analysis plan



The rest of the report provides detailed step-by-step guidance. The steps and codebooks described are designed with the aim of understanding:

- Whether and how specific groups of interest are represented in digital policies
- Whether the goals the policies set are linked to specific outcomes and relevant delivery mechanisms and stakeholders.

Understanding whether goals, delivery mechanisms and outcomes are aligned for the group of interest is important as this makes successful implementation and compliance with policy promises and commitments more likely.

Two theoretical frameworks underpin the methodology: the Social Policy Goal (SPG) framework[4] and the socio-digital inequalities framework.[5] In the relevant sections in this report, we repeat some of the information about the theoretical underpinnings as

[4] Liu et al. (2025).
[5] Helsper (2021).

discussed in the accompanying report, *Left out and misunderstood: Children in digital policies. A global review*,[6] because this is needed to understand how to approach the policies selected.

The DFC project that this toolkit is based on (see 'What does this toolkit do?') focused on representations of children (<18 years old), and the outcomes of interest were child rights in a digital age and wellbeing aligned with the Sustainable Development Goals (SDGs). However, depending on your group of interest the policy outcomes, stakeholders and delivery mechanisms that you want to evaluate or achieve may vary.

The rest of this report describes the different steps, as identified in Figure 1, and introduces the frameworks when relevant for analysis.

Step 1: Corpus construction

Corpus construction is the systematic process of identifying, collecting and organising a comprehensive set of documents that will form the empirical starting point for your analysis.

The corpus should be **representative** of the policies produced by the governing body you are interested in (e.g., an intergovernmental organisation [IGO], national government).

This step involves:

- deciding what the relevant policies are,
- determining the scope and boundaries of your search,
- establishing clear inclusion and exclusion criteria and
- implementing systematic search and collection procedures across multiple sources such as government databases, legislative archives and institutional repositories.

a) Define and contextualise the objects of study

Defining and contextualising the objects of study (e.g., digital policy or children) is essential to establish clear analytical boundaries, preventing incorrect and universalising findings.

While 'defining' objects of study creates categories to say what should and shouldn't be included, contextualisation allows more nuance. This means that definitions might shift depending on the unique legal, political and cultural realities of your study's setting.

[6] Helsper et al. (2025a).

For instance, in the DFC project, contextualisation meant recognising that cultural concepts of childhood and parental authority influence what *types* of policies refer to children and digital technologies in a given setting and incorporating that within the criteria for inclusion in the corpus.

To be able to select policies you first need to determine what you want to include and exclude as part of your corpus.

Defining and contextualising digital policy

The first step is **defining what counts as a digital policy,** as per the context and scale of your analysis. The specific definition of and inclusion criteria for policies may vary depending on the scale of analysis and the international, organisational/institutional, regional, national or local levels of digital governance.

For the purposes of this toolkit, we define policy as a document authored by a government entity, either a ministry or other official government institution, for which the accountability for its implementation lies with a governing body.

In the DFC project, this definition was modified and adapted to different country contexts based on their policymaking and governance process. Various documents encompassing legislation, bills, programmes, schemes and agendas were included under the umbrella of 'policy' when appropriate, considering the context of the country.

If you are analysing outputs from **intergovernmental or regional organisations,** the process of specifying what counts as policy may look different because:

- Many organisations publish retrospective research reports, trends and overviews, country and region profiles, or forward-looking frameworks, assessment tools, or guidance notes or explainers for policy practitioners, industry actors or non-governmental organisations (NGOs).
- Directives, agendas and declarations aimed at governments are scarcer and their influence on national policy is often indirect, achieved through promoting normative frameworks, provision of benchmarking tools and comparative data that governments can adapt to their specific needs.
- Treaties signed by member states are more likely to have an impact if there is a monitoring mechanism associated with these.

In this toolkit, we differentiate between digital transformation and digital inclusion policies:

- **Digital transformation policies** aim to develop new digital capabilities for the future (including infrastructure, platforms and human resources), and are often

led by ministries of infrastructure, economy, employment, education and defence/security.

- **Digital inclusion policies** focus on supporting certain areas, individuals or groups within the region or country who are at risk of being left behind or excluded from full participation in digital societies (including lack of access to: good employment, services, education, information, civic participation, a valuable social life, etc.). They suggest digital solutions to these social issues. They are often led by ministries of infrastructure, work, pensions, social security, education and health.

There is not always a clear separation between digital transformation and digital inclusion policies, although the first tend to be optimistic and future-oriented and the latter more sceptical and conscious of historical factors shaping current societies.

Top tip: Another consideration while defining 'policy' in the context of your project is considering where **authorship** and **accountability** for the policy lie.

In the DFC project, we decided that to be included on the national level, the policy must have been authored by a government entity, either a ministry or other official government institution. While these policies may be implemented alongside an NGO or private partners, the accountability for policy impact must lie with the government body.

Defining and contextualising your group of interest

The definition of the group of humans your study is focused on functions as both an **analytical lens** and a **boundary-setting mechanism**, determining the key frameworks that your analysis will be based on, what counts as meaningful representation or exclusion, and which policies to select for analysis.

The process of defining and contextualising this involves:

- the formal definitional boundaries (such as age ranges, legal status or identity categories) and
- the keywords that constitute direct and indirect references and representations in policy text.

This process can highlight the intersections between your target demographic and other identity markers that may form an important part of your analysis.

Example: You would like to focus on refugees in policy. You need to decide whether you will include policies that refer to asylum seekers, immigrants and undocumented people.

For the DFC project, we followed the United Nations Convention on the Rights of the Child (UNCRC) definition of 'child' as anyone under the age of 18. This decision was important because other organisations, such as the OECD, have a much broader definition, including what we would call 'young adults'. Most mentions of children found in digital policies are straightforward, and refer to 'children' as a one group, not specifying children's ages and developmental stages. However, some policies use more ambiguous terms, such as 'young people' or 'youth', which may refer to individuals over the age of 18. In such cases, contextual information was used to conclude if the policy refers to children or not.

- **Example 1**: 'Youth' is mentioned in the context of university education. This suggests that the policy refers to people over the age of 18 (with the exceptions of some students who start university before the age of 18, or in certain education systems where higher education may start earlier).

- **Example 2**: Mentions of young people in the context of financial inclusion or digital labour. In such cases, it is important to consider wider contextual realities and regulation of child labour in the chosen country/region.

To decide which policies to select for analysis and understand whether your group of interest is meaningfully considered, it is important to identify both **direct** and **indirect mentions** in your construction of the corpus. In the DFC project, policies referred to children directly, through mentions of terms such as children, infants, minors, adolescents, students, girls and boys, or indirectly, through mentions of parents, schools, teachers, children's social services, support workers, etc.

b) Select policies

In the selection of policies stage, you will:

- identify where to find policies,
- perform a search,
- determine the period for which policies will be included, and
- store the policies you have decided to include to create a corpus that you can analyse.

Identifying relevant databases, repositories and archives

When selecting policies, you have to have insights into where policies are likely to be found. This step is often **difficult and takes up much more time than you might expect.** Depending on the country/organisation, this might include:

- An official repository that is easily accessible and searchable. This is the easiest to search, but this is not readily available for all countries.
- Identify which ministries or departments are likely to publish and be responsible for policies relevant to your research and search and browse pages such as the 'Publications' or 'Legal instruments' pages.
- 'Press' pages of governmental agencies.
- Communiques, press releases or discussions of policy agreements that were made in person are diffused on websites or through news items. In countries where it is very difficult to find official documents, this is sometimes the only available information.

Top tip: For all context, a general search on an internet search engine combining at the very least the keywords 'digital' and 'policy' and the name of the country/region is important to make sure you have not missed anything.

Identifying and contextualising keywords and applying filters

Keywords used to find policies should be selected based on:

- A review of the literature.
- An initial review of readily available policies.
- Discussions with the research team after initial rounds of searches.

Contextualisation is important: **keywords** should be **adjusted for each country and governing body**.

The following keywords seemed to cover most policies in the countries and governing bodies included in the DFC project, and we think they will be useful for research in other contexts as well. In the end almost all searches included a combination of the following terms: 'digital'/'AI'/'ICT'/'cyber' + 'transformation', 'future', 'inclusion', 'divide', 'inequality', 'safety', 'security', 'economy', 'education', 'skills', 'literacy' and 'smart'.

The list of keywords should be translated and continuously updated as contextual understanding of the policy landscape strengthens.

The websites, databases or archives chosen may have different filters you can apply to aid you in your search for digital policies. **For example,** while building the corpus of policies from the World Bank, the DFC project used the preset theme filters 'digital technologies in education', 'digital transformation', 'cybersecurity', 'digital for climate' and 'broadband infrastructure, access and use'.

You may choose to exclude policies that are not associated with direct social outcomes, such as those primarily concerned with infrastructure and (mobile) spectrum awards, particularly if you aim to take a rights-based analytical approach. In the DFC project, we found that some digital policies mainly discuss technical specifications and licenses – especially cybersecurity policies, policies about national security against foreign actors or those focusing on spectrum and digital transactions – and they do not have socioeconomic or sociocultural goals.

Since we aimed to study the impact of digital policies at the intergovernmental level on regional and national policy discourses in the DFC project, we included guidelines aimed at governments and other governmental actors such as policymakers, the judiciary and information commissioners. This is because policies from such organisations are often not legally binding but enact a soft power on countries' digital agendas. However, we excluded those directed at non-governmental and corporate actors, such as practitioner notes, frameworks and assessment tools.

Determining the period of interest

You should **decide on a time limit** for inclusion of policies in the research and **justify the time frame chosen**.

For the DFC project, we decided to look at policies published (electronically or otherwise) between January 2020 and July 2025. However, older policies were included if they were still underpinning current policy and interventions, or were referred to in new policies or official communications from the government, with a maximum cut-off point of 10 years applied especially in countries with a limited number of policies. The period since 2020 was chosen purposively to include the onset of the Covid-19 pandemic, which saw increased attention to issues of digital exclusion as people came to rely heavily on ICTs.

Data management

It is important that you keep an offline archive (i.e., the actual corpus) and **download all policies**. A PDF or other available format should be **saved to a secure location**. If it is a website this can be done by printing the website in PDF format and saving this. This is an

important step, particularly as policy documents can be retracted (e.g., during political instability) or websites can become inaccessible.

It is also important to **record the date of download,** in case the text published is edited during or after your analysis, so that the credibility of your analysis is not compromised. Create a separate document where you have an overview of all policies collected, in which you detail the name of the policy, the ministry responsible, the date published and a link to where you found it.

c) Identify challenges

It is important to note **which challenges were encountered** when constructing the corpus and **how decisions were made about inclusion or exclusion** of policies. This should be reported on for each region or country analysed, alongside the strategies used to mitigate these challenges. Here we describe a few issues we encountered for the DFC project, which we think you are also likely to encounter, and how we addressed them.

Difficulty in defining 'policy'

Determining what a policy is can be difficult, particularly if conducting multinational analysis, and especially trying to make a distinction between regulation and legislation, policy and law. This includes differentiating between regulations that lay down the procedure to do things, like acts and bills relevant to stakeholders in the digital realm, such as platforms and constitutional instruments (e.g., general [non-digital] conventions, treaties, declarations and deciding which types of documents are relevant for your research questions). This varies between contexts, as we learned from our research. Below you will find a description of some of the challenges encountered and decisions made about inclusion or exclusion from the corpus in the DFC project.

- Often countries have white papers in which governments float ideas and consult stakeholders before implementing them in policies. While interesting in terms of the policy process, we did not include them because policies are adjusted or are dropped altogether, and because they are sometimes used to test the water for ideas that are too controversial and never seriously on the agenda.
- Many countries, such as Indonesia and Vietnam, mostly have legal instruments. We decided to include legal instruments for these particular cases.
- In countries such as India, where digital regulation is in the early stages, especially around issues such as data protection and privacy, legal instruments are developed before policy initiatives, and offer key insights into the discourse surrounding these issues. We decided to include some legal instruments in these particular cases.

- Policy in the USA is defined as all actions taken by the federal government; it only has legal instruments on the national level, while 'policy', as we defined it, could be created by state governments.7 We decided to analyse policies of states that exemplified different approaches to digital transformation and inclusion.

Policies beyond the digital

As far as possible, only digital transformation and digital inclusion policies were included. **Policies that were not explicitly digital** were only included if they had a substantial section on 'digital', particularly education policies.

Some national economic or education plans do not have a dedicated section on 'digital' but have multiple references to digital aspects scattered throughout. **Decisions** about whether to include such policies can be made **based on the significance of the policy in the policy landscape, or its impacts on industrial and social sectors**. If the policy plays a major role in setting the agendas for other digital-specific policies or has serious implications for the operations of the technology sector and/or impacts on the digital services and experiences of the demographic you are analysing, then such policies can be included.

Policies not directly relevant to the group

When analysing policies in the context of the DFC project, children were logically not expected to be included in policies that only pertained to adults without a direct impact on them, such as policies about the adoption of ICTs in business or government functioning (e.g., e-government services, banks), those related to employment (with the consideration of the reality of child labour in many countries), entrepreneurs and farmers, as well as policies focussed on pensioners.

We did, however, include policies focusing on the 'economy' as a whole in terms of the future of work and necessary skills, and those offering guidelines for economic organisations about content and design of digital platforms and non-governmental services in a digital world, whenever these might impact children's lived realities. We excluded policies related to retirement and pensions and financial management (e.g., online banking, micro-payments).

Reach or scope of the governing body's responsibility

For international and intergovernmental organisations, there is a decision to be made about whether country-specific reports should be included or if they should be included in the country policy analysis. For the DFC project, we decided to include some country

[7] Helsper et al. (2025b).

reports in the national policy analysis only if they were directly referred to by the national government as guiding policy. They were excluded from the IGO policy corpus.

Access and persistence issues

Accessibility of policy texts varies widely between countries, and depends on various factors such as the degree of centralisation of the policy landscape, levels of bureaucracy and/or bureaucratic cultures, levels of digitisation of policy documents and general policy and documentation processes, as well as political (e.g., autocratic or democratic regimes), geographical, and infrastructural limitations.

For the DFC project, we found that in some countries, such as Niger, Senegal and Togo, there was limited access to full policy texts, with some documents available only as summaries or in news reports. We also faced difficulty locating official government sources and at times publication dates for the policies. For Mozambique, policies were found in databases or official government announcements mentioning them, but without direct download links. Some government websites were frequently down, were inaccessible from the location the researchers were based in (i.e., the UK) or required the researchers to provide ID and give personal data. In many countries there were public consultations, campaigns or programmes that were not anchored in a specific policy. In the case of Angola, a strategy or agenda was sometimes mentioned in press articles or consultancy reports, but no official document could be found.

There is also inconsistency in how digital policies are recorded and structured, with accountability shared by multiple ministries, often without a central repository, as noticed in the cases of Senegal, Togo and India. For a policy from Togo, there was no formal policy document, but reports on implementation were available. For Mozambique, databases were the best source for the policies, but without premium access, only the policy name, number and a brief description was visible. In the case of the USA, initial searches were found to be overwhelming, with lots of US-affiliated non-profit and state organisations posting about digital inequalities, but few centrally commissioned and government-owned policy documents.

Another issue was the disappearance of policies even if they were still referred to as relevant to current policymaking and interventions. In the case of India, we had access to policies through a previous project that we could no longer locate under the link where we had found them previously or elsewhere.

Step 2: Basic analysis

Following the construction of the corpus, the policies will then be broadly coded to identify and categorise:

- References to the group of interest (e.g., children)
- General mentions of inequalities
- Overarching policy discourses (see Step 2(b) for different types).

A discourse refers to the political-economic intentions and ideologies that underpin the policy content and goals.

A general basic codebook was designed for the DFC project to understand in which context children were mentioned and which policy types they were generally not considered in (see Table 1). This codebook helps identify policies that will be analysed in-depth in terms of how they represent children and what mechanisms, if any, they propose to make sure children of all backgrounds are included in a digital present and future.

The codebook includes the general coding category, general codes within that category with options to be selected and a detailed code, which consists of examples of potential words or expressions used that lead to this general code. The latter are entered in the original language in the 'Notes and quotes' column.

a) Identify mentions of the group of interest

In the basic codebook, a distinction is made between direct and indirect mentions of the group of interest.

Direct mentions are mentions of the group as a whole or its members, and **indirect mentions** are mentions of services, institutions, professionals and other individuals directly in contact with the group (examples of these for the DFC project are given in Table 1).

For other groups these will be different – in the case of immigrants, for example, direct mentions may include mentions of migrants, immigrants, refugees, expatriates and asylum seekers, indirect mentions may include mentions of border forces, immigration lawyers, asylum case workers and visa sponsors.

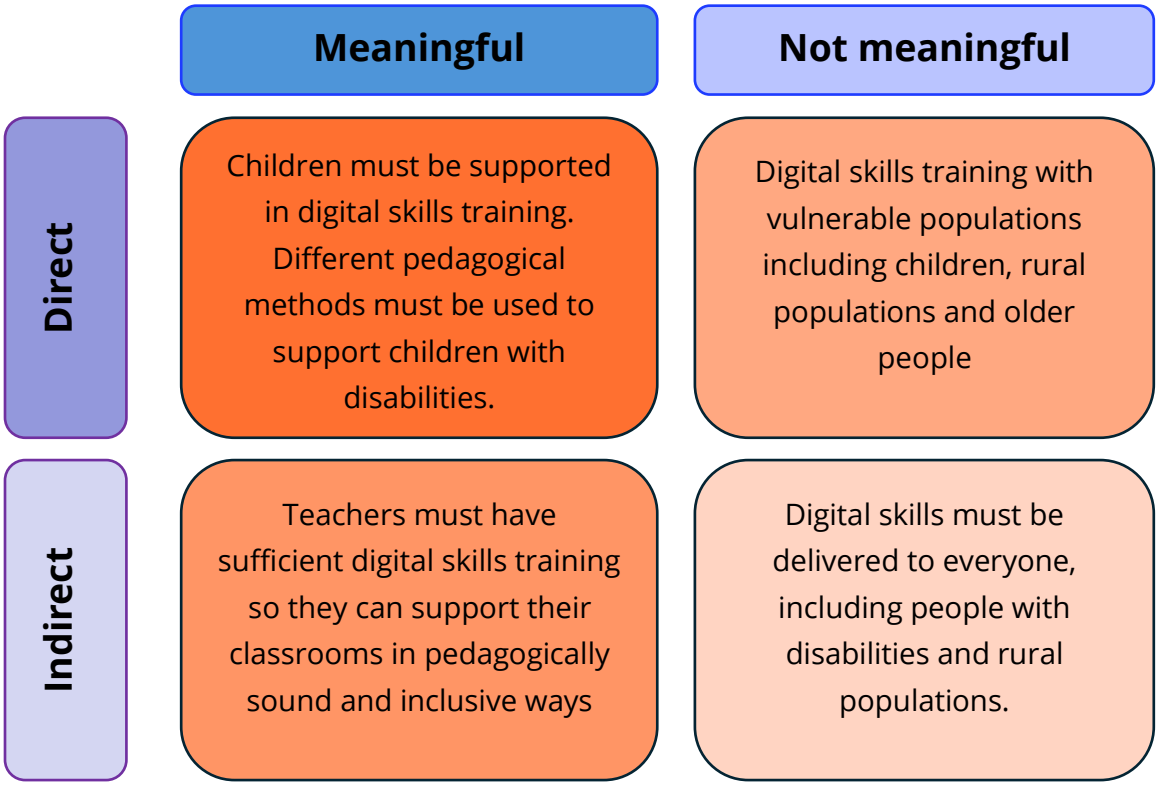
Further, the mentions should be coded as **meaningful** or **not meaningful**. This code determines whether a policy should be taken to the next step of detailed analysis. For the DFC project, we defined meaningful mentions as mentions of children that are not just as part of a list of vulnerable populations. This is because we found that many policies do

mention children, and that if they do, it is often in a cursory way, as part of a list of vulnerable populations, without specific engagement with the realities of children from different backgrounds, even if policies directly impact them.

Table 1: Basic codebook (for the DFC project)

Coding category	General code	Subcategory code (DFC project examples)	Notes and quotes
Mentions children	No, not mentioned at all Yes, mentioned directly Yes, mentioned indirectly Yes, meaningfully mentioned (more than as part of a list)	Directly (e.g., youth, girls, boys, teenagers, toddlers, children, young people, etc.) Indirectly (e.g., parents, mothers, fathers, students*, teachers, disadvantaged households/schools; children's services, etc. * Since 'students' mostly refers to university students, this is recorded as an indirect mention, but can also be classified as a direct mention depending on the language and country context	Give exact terms used in original language
Policy discourse and goals	Economic development	Economic prosperity and growth for the economy and workers	
	Social inclusion	Promoting social equity and justice for disadvantaged or marginalised groups	
	Civic participation	Increased civic engagement and responsibility (nation building) for the benefit of democracy, society and citizens	
	Human rights	Enhanced opportunities for individual development and wellbeing – according to the Universal Declaration of Human Rights (UDHR)	
	Child rights	Promoting, respecting, protecting and fulfilling all children's rights in the digital environment – according to the UNCRC and General Comment No. 25	
Mentions inequalities	Yes/no	Race, class, caste, ethnicity, gender, sexuality, sexual orientation, disability, wealth/poverty, education status, employment status, rural/urban, migration, age, religion, etc.	

Figure 2: DFC project example of coding direct/indirect and meaningful/cursory mentions of children



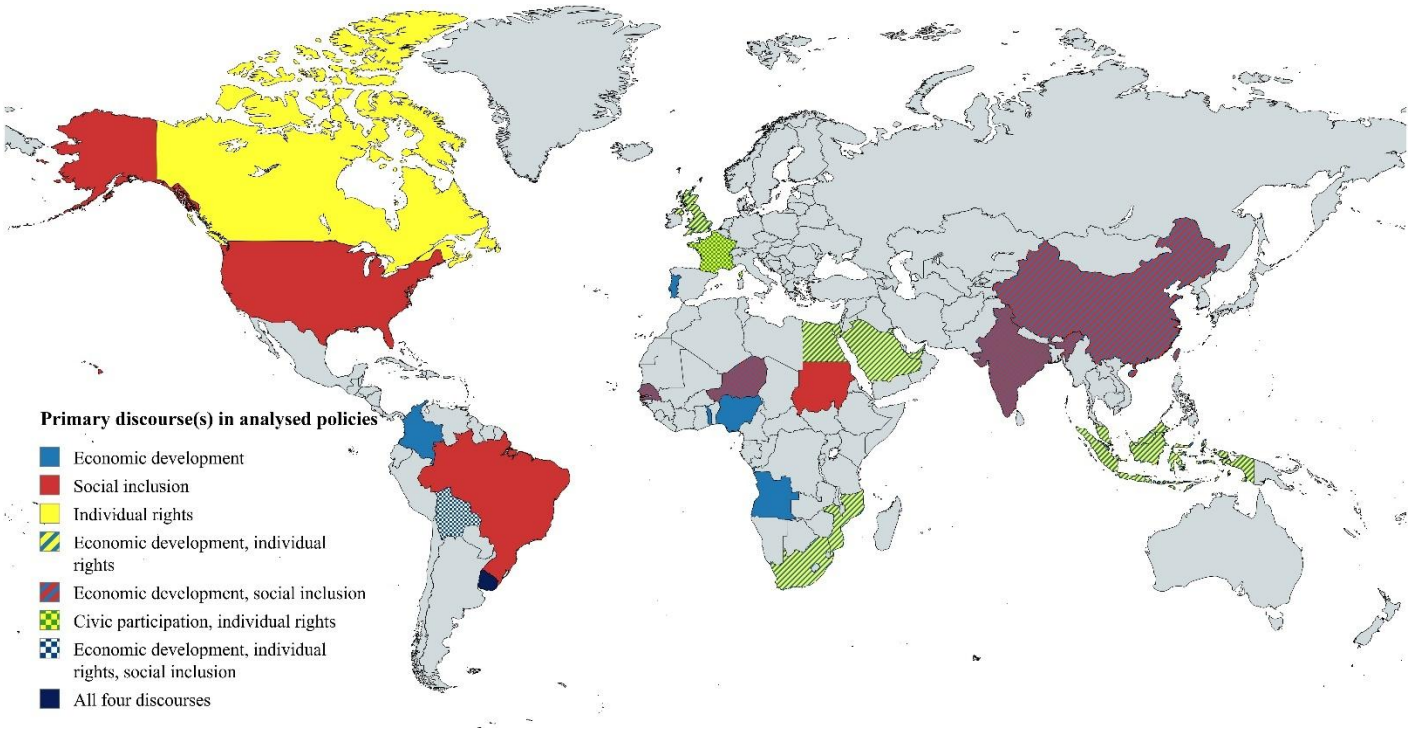
b) Identify policy discourses

Once you have selected and downloaded the policies that mention your group of interest, the next step is to categorise which type of policy has mentioned your group the most. Here the Social Policy Goal (SPG) framework[8] can be used to theorise which types of digital policies exist. The SPG framework takes a policy-as-discourse perspective and assumes that political-economic circumstances and governing ideologies shape policy content and implementation, reflecting struggles over how (national or regional) reality is to be perceived. More information on the framework and how it was applied in the DFC project can be found in the report *Left out and misunderstood: Children in digital policies. A global review.*[9]

It is important to note that various discourses can be present at the same time and should be coded as such. For an example of how this might be visualised, see Figure 3.

[8] Liu et al. (2024).
[9] Helsper et al. (2025a).

Figure 3: Map of dominant policy discourses around the world (DFC project)



Note: While the most dominant approaches are indicated on the map, in practice, separating discourses was difficult – different policies tackle different discourses within each country.

To support the identification of mentions of children and policy discourses, the policies should be coded for **mentions of inequalities**. Here, you may code for mentions of inequalities generally in the policy, as this serves to situate the detailed analysis, and because groups of interest are included in other groups. In the DFC project, children are also part of groups such as people with disabilities, refugees and those living in poverty, and these can therefore be considered indirect mentions. However, this is potentially time-consuming.

Tip: If you are constrained by resources, you may choose to code for mentions of inequalities only in reference to the group you are analysing. For example, in the DFC project, inequalities with respect to children included mentions of gender inequality (girls are mentioned), disabled children and low-income households

Improving the codebook

The continuous and iterative development of the codebooks, and detailed notes from the DFC project on which this toolkit is based, have led to the above version of the codebook which should be fairly generalisable to the analysis of other groups of interest.

Nevertheless, there are **likely to be adaptations needed** and issues encountered even if you are also interested in representations of children. This is particularly likely if you are conducting multinational or multiscalar research and/or working with a large team of researchers. This may be caused by researchers' subjective interpretations of the policy discourses, use of contradictory ideologies and/or goals in policy text or general funding and time limitations.

Appropriate strategies to avoid this **are frequent team meetings** to discuss selection and coding, conducting peer reviews or, in the case of limited resources, choosing only a select number of the most important codes for the research objectives.

Most importantly**, take extensive notes** about the rationales behind the choices made for each policy context.

c) Select key policies for in-depth analysis

If your research is focused only on finding how many or which policies mention the group meaningfully, you may end your analysis at this stage. If not, take the policies that have meaningful mentions of the group of interest for the next step: in-depth analysis.

The DFC project excluded child-only policies in most cases at the first step of basic analysis because this step was about identifying whether broader policies mentioned children and in which way. It was logical that they would be mentioned in child-specific policies. In countries with limited policy availability, child-specific policies were included at the next step, in the detailed analysis, because they were the only place where children were mentioned in relation to the digital.

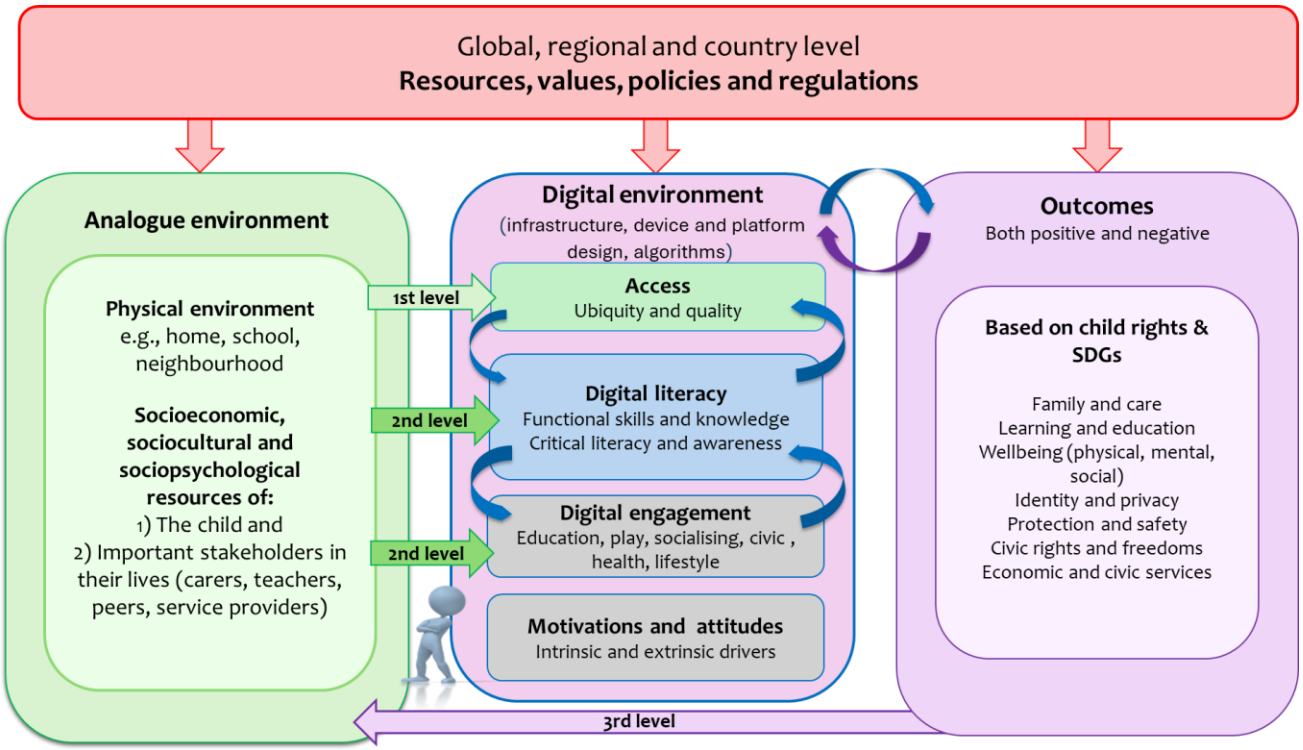
Step 3: In-depth analysis

The last step before reporting and dissemination is also very time-consuming; it involves a deep dive into the selected policies, which requires a real understanding of the policy landscape and the forces that shape it within the context you are studying. The first step is to see whether the mentions of your group of interest have goals, delivery mechanisms and outcomes that are aligned, followed by a qualitative analysis of how your group of interest is represented and involved in the policymaking process.

a) Analyse goals, delivery mechanisms and outcomes for groups of interest (and how they align)

The detailed analysis of digital policies developed for the DFC project is based on the framework first applied by the Digital Skills to Tangible Outcomes (DiSTO) project[10] and further developed by Helsper (2021) into the socio-digital inequalities model, which was adapted with relation to children for the DFC project (see Figure 4).

Figure 4: Framework for the analysis of digital policy mechanisms and outcomes in relation to children



It describes how global, regional and country resources and policies shape the analogue and digital environment, and how this leads to differential outcomes from digitisation.

In the in-depth coding phase, this framework guides the analysis of the representation of different **beneficiaries** and **stakeholders** (i.e., the analogue environment), different digital **delivery mechanisms** (i.e., aspects of the digital environment) and **goals** identified by the policies (i.e., outcomes). In the DFC project, the framework was adapted for children by identifying physical environments and socioeconomic, sociocultural and sociopsychological resources relevant to children.

[10] From Digital Skills to Tangible Outcomes (DiSTO): www.lse.ac.uk/media-and-communications/research/research-projects/disto

The **DFC study** based the framing of outcomes on the UNCRC,[11] General comment No. 25[12] and Sustainable Development Goals (SDGs).[13] For your study, other policy outcomes and compliance standards might be more appropriate.

Example 1: For projects focusing on refugees or Indigenous populations, the Universal Declaration of Human Rights (UDHR)[14] or the 1951 Refugee Convention[15] may be more appropriate

Example 2: For projects focusing on representations of those experiencing economic and/or labour-related precarity, the Multidimensional Inequality Framework[16] could be helpful to determine which outcomes policies should refer to.

The codebook should be adapted to your research topic, particularly the detailed codes for stakeholders, beneficiaries, identified analogue problems and aimed-for outcomes. These should be coded to indicate the presence or absence of each of the elements in policies, with additional qualitative observation and quotes.

Direct quotes from the policy texts are especially important. This allows you to keep record of the actual language used, to understand the context and the specific meaning of the references to certain aspects and groups of interest.

These quotes can be used later to illustrate important points in your analysis. This is key data that can be shared and discussed between researchers and used for future publications. Sometimes you only figure out at the end that something is important, and reading through the codebooks at various stages in the project can give new insights (you do not want to have to read through all policies again)!

The codebook is applied to each selected policy individually and should always include the following:

1. **Stakeholders** (e.g., parents, teachers, NGOs, tech companies, ministries) that are poised to deliver solutions or to be intermediaries in interventions

2. **Beneficiaries** (e.g., children, vulnerable groups, workers, businesses, schools)

[11] United Nations Convention on the Rights of the Child: www.unicef.org.uk/what-we-do/un-convention-child-rights
[12] United Nations Committee on the Rights of the Child, General Comment No. 25: www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation
[13] United Nations Sustainable Development Goals: https://sdgs.un.org/goals
[14] Universal Declaration of Human Rights: www.un.org/en/about-us/universal-declaration-of-human-rights
[15] The 1951 Refugee Convention: www.unhcr.org/about-unhcr/overview/1951-refugee-convention
[16] The Multidimensional Inequality Framework, London School of Economics and Political Science: https://sticerd.lse.ac.uk/inequality

3. **Analogue problems identified** – related to socioeconomic (e.g., poverty, education), sociocultural (e.g., civic participation, belonging) and sociopsychological wellbeing (e.g., relationships, physical and mental health)

4. **Digital problems identified** – first-level (i.e., access, attitudes, biased business/algorithmic models), second-level digital resources (i.e., skills, engagement, content) and third-level inequalities in outcomes (i.e., experiences in the digital environment) of the group of interest

5. **Delivery mechanisms to tackle the problems** (e.g., legislation, provision of infrastructure, skills training, content/services) related to the group of interest

6. **Predicted outcomes of digital interventions proposed** (e.g., economic development, civic participation, wellbeing, human/child rights, SDGs), including any mention of Key Performance Indicators (KPIs) specific to your target group.

In what follows, we elaborate on the sections of the advanced codebook corresponding to these different socio-digital categories, as developed for the DFC project.

If analysing different countries or governing bodies, separate documents should be created for each. The overall codebook for a specific governing body or country for each policy is created by putting all the different tables together, as presented in the toolkit. For the DFC codebook we used Excel, and the codebook can be recreated by copy-pasting the different tables in this toolkit underneath each other for each policy.

In the codebook corresponding to a particular governing body, each policy had its own section providing a link to the policy, its name, its year of publication and date of accessing the link and downloading the document. In the tables presented in this toolkit, underneath this section the rows correspond to the socio-digital category, overarching category and detailed subcategory codes. Table 2 illustrates what this could look like in a newly created codebook.

Table 2: Illustration of how codebooks can be organised

Name of policy			
Link to policy			
Year of publication			
Date accessed/ downloaded			
Mentions children	(Write down terminology used)		
Socio-digital inclusion category (see below)	**Overarching category code** (main categories to be coded within the socio-digital inclusion category)	**Detailed subcategory code** (categories within the overarching coding category – one row for each code)	**Notes and quotes** (in the original language and translated)
Stakeholders			
Beneficiaries			
Infrastructure and access			
Attitudinal drivers			
Literacy and skills			
Engagement and content			
Outcomes			

Note. The DFC codebook had a separate Excel spreadsheet for stakeholders (1) and beneficiaries (2) and for the other socio-digital environment and inequalities indicators (3–6).

1. Global, national/regional and country-level stakeholders

In this section, the policies should be coded for mentions of global, regional or national governmental, funding and implementation bodies, alongside other local stakeholders involved in delivering the digital interventions (directed towards children).

Detailed **qualitative notes** should be provided on the **role played** by each type of stakeholder, and these should be **accompanied by quotes** from the policies that **illustrate these observations.** Table 3 shows how these were coded for stakeholders at the global, regional/national and local level in the DFC project.

Table 3: Codes for global, national/regional and local stakeholders

Overarching category code	Detailed subcategory code (DFC project examples)	Notes and quotes
Global ecosystem		
Intergovernmental	The World Bank	
	UN (e.g., ITU, UNICEF, UNESCO)	
	OECD	
	WHO	
	Regional bodies (e.g., European Union, African Union, ASEAN, MERCOSUR)	
	Other	
(I)NGO	Human rights (not child-related)	
	Educational/learning	
	Child rights	
	Women's (rights)	
	Digital inclusion/literacy	
	Other	
Platforms/ICT industry	Tech companies providing devices/services/training	
	Tech companies providing content/infrastructure/ networks/security	
National/regional ecosystem		
Government ministries	Transport/infrastructure	
	Economy/business	
	ICT/science	
	Education	
	Health	
	Culture and media	
	Other	
Industry and content regulation	Regulation content	
	Regulation algorithms	
Architects/infrastructure	Smart city engineers	
	School design/infrastructure	
	App developers	
Charities	Child-related	
	Family-related	
	Education-related	
	Health-related	
	Poverty/inclusion-related	
	Other	
Private sector	Telecoms providers	

Other national stakeholders		
Local ecosystem (in the immediate analogue environment, e.g., neighbourhood)		
Governance	Local and community government/council	
	Mayoral office (equivalent)	
	Social work and other support services	
Schools	Teachers	
Libraries		
Households	Parents, family	
Religious organisations	Churches, mosques, synagogues, temples	
(Youth) clubs	Sports, arts and crafts, Scouts, etc.	
Community centres		
Telecentres, IT cafes		
Civil society	Local NGOs, community activist organisations	

2. Beneficiaries and 3. Analogue problems

Table 4: Codes for beneficiaries

Overarching category code	**Detailed subcategory code** (DFC project examples)	**Notes and quotes** (include term used in the original language and translated to English)
Economic	Rural/urban/region/state	
	Wealth/poverty	
	Education	
	Employment	
	Class	
Identity	Caste	
	Race/ethnicity	
	Gender, sexuality	
	Religion	
	Migrants/citizenship	
Capacity/care	Disability	
	Age	
	Orphans or 'left behind children'	
Other		

Beneficiaries are **subgroups of the group of interest** who are considered differently vulnerable based on the Socioeconomic, sociocultural and sociopsychological resources of

the groups they belong to and the environments in which they live their everyday lives. The problems they are said to suffer from in the policy (e.g., poverty, lack of education, ill health) should be noted in a qualitative manner in the codebook. This involves coding identity aspects, such as the groups (e.g., race, gender, religion, ability) they are part of; the socioeconomic status of their households; the services and activities provided to them in their neighbourhoods; and the material conditions of their workplaces or schools.

Notes should be provided that clarify which assumptions are made about the subgroup and the **problems they face,** and if **any causes** of these problems are **identified**, quotes should be given to illustrate these points. In the DFC project, the codes given related to subgroups of children that could be considered vulnerable, such as orphans, religious or minority ethnic groups, children in rural areas and LGBTQ+ youth. Categories were added for specific countries when the general categories did not apply (see Table 4).

4. Digital problems identified and 5. Delivery mechanisms to tackle the problems

This part of the codebook identifies which **digital resources** people (and specific subgroups of them) have as related to the digital environment. There are four subsections to the digital problems and associated delivery mechanisms identified in policies, which are related to:

- **First-level digital inequalities**, separated into (1) infrastructure and access and (2) attitudes and motivations indicators
- **Second-level digital inequalities**, separated into (3) skills and literacy and (4) engagement and content (i.e., online experiences).

For all these codes, **KPIs**, if given, should be included with a **note on whether these are specific to the group of interest and whether subgroups are mentioned.** The KPIs in this section should be related to experiences and outcomes to be achieved in relation to their digital environment, and not the economic, social, civic or wellbeing outcomes. These KPIs[17] are important because it is what governing bodies should be held accountable to. There are many problems with KPIs, but their absence is even more problematic.

Infrastructure and access (first-level digital delivery mechanisms and inequalities)

This category looks at what **infrastructure and devices** are referred to as being insufficient or unequally distributed in the policies. These can be mapped on to the distinction between quality (e.g., higher speed, greater bandwidth, more features), ubiquity

[17] Quantifiable metrics used to track and evaluate the effectiveness and impact of an organisation, project or initiative, in this case, a digital policy.

(e.g., data plans, public Wi-Fi, mobile connectivity) and autonomy (e.g., ownership of device, no monitoring of access or use, data protection and control over data) in the notes.

In the **notes** for these codes there should be a description of **where and for whom infrastructure and access is a problem**, what **interventions** involving infrastructure or access provision will reach, and **what this is about**, for example, lowering costs, providing access or regulating unfair practices based on algorithmic bias and structuring (see Table 5). Full quotes from policies should be included to illustrate the points made. When it comes to children, providing internet access or devices to schools in rural areas is an often-encountered **delivery mechanism** for digital policies aiming to tackle infrastructure or access problems. Less common in the DFC project, but increasingly part of the discourse, is regulation regarding data protection and algorithmic modelling that limits or pushes access to certain platforms, services and content in unequal ways.

Table 5: Codes for infrastructure and access

Overarching category code	Detailed subcategory code	Notes and quotes
Infrastructure (region, neighbourhood, school)		
Established	Fibre, broadband	
	Mobile connectivity, 5G	
	Satellite	
	Innovation hubs, tech parks	
Emerging (to be updated according to the latest trends)	Cloud and data infrastructure (data centres, cloud platforms, etc.)	
	Internet of Things, Virtual Reality, Augmented Reality, artificial intelligence	
	Smart cities and infrastructure (smart cities, smart energy grids/meters, smart transportation etc.)	
Access (individual/household/ownership)		
Connectivity	Data plans	
	Broadband/Wi-Fi plans	
Devices	(Smart) phones	
	Other portable devices (tablets, laptops)	
	Fixed PCs	
Next generation	IoT, VR, headsets	
Accessibility	Screen readers, large letters, etc.	
Autonomy	Control over access to and use of personal data	
	Algorithmic selection in presenting apps, platforms and content	

Attitudinal drivers (first-level digital delivery mechanisms and inequalities)

Attitudinal drivers are consistently included in theories around digital inequalities as one of the factors determining how digitally included someone is or feels. Despite this, specific conceptualisations of the different types of attitudinal drivers, such as the motivation to use technology vs attitudes towards technology, is less present. To remedy this conceptualisation gap, for the DFC project a distinction was made between individual motivation (intrinsic drivers) and general awareness (extrinsic drivers) of the benefits and risks around technology (see Table 6).

A **delivery mechanism** to create more constructive attitudes might be outreach to community organisations with the aim of increasing interest in using technologies for specific purposes in the group of interest, public awareness campaigns to create positive attitudes about the wider benefits of digitisation for society as a whole, or the opposite, creating awareness of people's rights to disconnect or protect their data.

In the DFC project, some of the intrinsic motivators that might be mentioned as needing to be stimulated were how digital technologies might improve a child's quality of life and supporting learning for those with disabilities, while extrinsic drivers might be greater opportunities for employment and employability, awareness of risks related to data protection, privacy or bias.

For these codes, make notes and collect quotes about **normative assumptions about intrinsic factors driving engagement with digital technologies** (e.g., 'technology is not for girls'). For extrinsic factors, code for **how proposed technologies are expected to be beneficial and which risks are to be mitigated** for the group of interest.

Table 6: Codes for attitudinal drivers

Overarching category code	Detailed subcategory code (DFC project examples)	Notes and quotes
Intrinsic motivation/ individual interests	Participation (civic)	
	Professional future (including education)	
	Personal wellbeing	
General awareness/ extrinsic motivation	Benefits	
	Risks (safety, data protection, bullying)	

Literacy and skills (second-level digital delivery mechanisms and inequalities)

In the literature several different types of **skills and literacy** are identified:

- Self-confidence (often linked to motivation)

- Technical skills, ranging from the basic skills needed to use and operate devices and websites to more advanced technical programming skills
- Information skills indicating an ability to navigate and evaluate the trustworthiness of information
- Interactional skills that help to engender positive communication and interactions with others
- Content production skills to create engaging content and reach a wide audience.

All these have a **functional aspect** (being able to use the features and functions of technologies) and a **critical element** (understanding how technologies work, why they are designed the way they are designed and how they can have varying impacts on different people).

Notes in the codebook should provide descriptions and quotes (where possible) of **where and for whom a specific type of digital skill** or literacy **is a problem** and whether the digital skills training is **benefit and opportunity or risk and safety oriented**. Notes should include a description of whether the skills and training described mostly address **technical, functional skills** or relate to **knowledge and critical awareness** of why certain content is produced and how algorithms might be biased.

Table 7: Codes for literacy and skills

Overarching category code	Detailed subcategory code	Notes and quotes
Skill/literacy type		
General mention	Self-confidence	
Dimension	Advanced/programming	
	Technical/operational	
	Information navigation and evaluation	
	Communication and interaction	
	Creation and production of content	
Aspect	Functional and instrumental skills (using technology expertly)	
	Critical understanding and knowledge (understanding how technologies work and the impact they might have)	
Delivery mechanisms (DFC examples)		
Formal	Digital curriculum or certification	
Informal	Autonomous learning, information on social media	

Note: Skills and literacy categories tend to be universal for different groups of interest; delivery mechanisms might differ.

Notes should also describe the **delivery mechanisms** (curriculum design, formal, informal education) in more detail, and whether this is for teachers, children or parents. In this part of the codebook, delivery mechanisms are included as a separate category to code (see Table 7).

An example of a delivery mechanism from the DFC project is the rollout of a digital skills curriculum focused on coding as a technical skill targeting youth not in education, employment or training (NEET) who are assumed to lack digital literacy.

Engagement, availability and experience of content and services (second- and third-level digital delivery mechanisms and inequalities)

Table 8: Codes for engagement and availability of content and services

Overarching category code	Detailed subcategory code	Notes and quotes
Learning	Virtual or online learning platforms	
	Educational content	
	Appropriate information/news	
Financial	Employment	
	Micro-financing/payments	
	Mobile banking	
	Cybersafety (including scams and employment rights of the child)	
Civic	Healthcare	
	Social services	
	Democratic participation (voting, assemblies, town halls, etc.)	
	(e)Governance, digital ID	
Social/family life	(Appropriate) social media platforms	
	Family/parenting support	
	Online communities	
Leisure	(Appropriate) entertainment (non-educational TV, film, video, music, etc.)	
	Sports, hobbies and other activities	
	Content creation (active)	

There are also considerable differences in how people **engage with technologies and how they experience this engagement**, often linked to their sociocultural backgrounds (e.g., gender, religion, ethnicity, class or caste). This can relate to the availability of content and its perceived usefulness or attractiveness to particular groups, or to the mis- or underrepresentation of certain narratives and experiences due to algorithmic bias. For the DFC project, policies noted that there might be differences between boys and girls in how much they socialise with others online and in what kinds of games they play online. This

can be framed as a potential problem if it is considered to be inappropriate or less valuable if certain content is engaged with by (certain groups of) children (see Table 8).

Notes and quotes should indicate which **group of people is linked to engagement with certain content or lack thereof**, whether the content appropriate/important for these groups is **available online**, and how **experiences** of digital services or platforms might differ. Notes should also indicate what the policy proposes to do about this (i.e., the **delivery mechanism**). Online safety regulations that protect girls against technology-facilitated child sexual exploitation and abuse (CSEA) are an example of an intervention in this area, but so is the provision of educational content in languages spoken by different tribes or ethnic groups, or the regulation of algorithmic bias or misinformation.

6. Predicted outcomes of digital interventions proposed, including KPIs

Finally, there are the **outcomes of digitisation** and engagement with digital technologies, related to the third level of digital inequalities (see Table 9). Depending on your research questions, target audiences and/or funder requirements, you should choose appropriate underpinning frameworks to inform the outcomes section of the socio-digital inequalities framework.

Here, **detailed notes** should be provided on each code for the **expected/intended non-digital outcomes** from the policy. Record observations and quotes about **the key performance indicators (KPIs)** relevant to each detailed code, whether they are meant to **increase benefits or prevent harm, and for which groups**. While ideally the outcomes would be clearly stated and articulated in terms of precise KPIs for evaluation and accountability, often outcomes are implicitly referred to throughout the policy text through mentions of why the intervention will be significant and which international standards it meets, among others. Notes should clarify whether the **outcomes are explicitly connected to frameworks** referenced in the policy on which the codes are based (e.g., the SDGs), whether **directly or indirectly** address the outcome itself, or are actually descriptives of **digital outcomes rather than analogue ones**. Quotes are very important in this section in particular.

In the DFC project, the codes that were used varied, although there were many digital policies referring to reducing child poverty, violence against girls and women and improving education, health, freedom of expression, age-appropriateness, privacy and safety. A few referred to overcoming inequalities in learning and early childhood development, standard of living, awareness of children's rights and consultation. Notes and quotes should indicate whether there are clear KPIs associated with each of these.

Table 9: Codes for outcomes (third-level digital inequalities)

Overarching category code	Detailed subcategory code (DFC project examples)	Notes and quotes
SDG global indicators	Any SDG (not related to children, as in the list below)	
	Child poverty	
	Nutrition	
	Maternal mortality, child mortality and universal healthcare	
	Learning and early childhood development	
	Violence against girls and women and harmful practices	
	Drinking water and sanitation and hygiene	
	Child labour	
	Abuse, exploitation and violence	
	Birth registration	
Child rights	Relax and play	
	Freedom of expression	
	Be safe from violence	
	An education	
	Protection of Identity	
	Sufficient standard of living	
	Know their rights	
	Health and health services	
Digital child rights	**Equity and diversity:** Treating all children equally, fairly and supporting vulnerable children	
	Best interests: Children's best interests a primary consideration	
	Consultation: Children meaningfully consulted in policy or product development	
	Age appropriate: Appropriate for child users or adaptable for children of different ages	
	Responsible: Reviewing and complying with laws and policies relevant to children's rights	
	Participation: Enabling children to participate in digital publics	
	Privacy: Adopting privacy-by-design in policy and product development and use	
	Safety: Adopting safety-by-design in policy and product development and use	
	Wellbeing: Enhancing, not harming, children's mental or physical health and wellbeing	
	Development: Enabling children's learning, imagination, play or belonging	
	Agency: Taking steps to reduce compulsive and exploitative product features	
Other outcomes	Financial and material security (incl. employment, wealth)	
	Civic participation	
	Social capital	
	Family life	
	Privacy and data protection	
	Discrimination and harassment	

After you have completed the coding of this final stage you should **summarise the general findings across all the coding categories**, and identify whether all the different elements are lined up by asking:

- **Are specific problems identified** for your group of interest and subgroups within these?
- **Are clear digital delivery mechanisms identified** that aim to tackle the problem (with stakeholders identified that can support this implementation)?
- **Are KPIs identified** in relation to your (sub)group of interest that will indicate whether the policy was implemented successfully?

Figure 5: Visual of the (mis)alignment between problems identified, delivery mechanisms and KPIs for the issue of poorer children in rural areas being less likely to go to school

Problem statement	Are causes of the problem identified?	Is there a delivery mechanism?	Is there an associated KPI?	Do they align?
Poorer children do not have access to schools.	Yes. Policies will provide connectivity and online learning programmes.	Yes. The rich-poor gap in participation in formal education will decrease by 30% and poor children's earning outcomes will improve by 0.5 after policy implementation.	Yes. There is a logical progression and alignment between beneficiaries, causes of the problem, delivery mechanisms and KPIs.
Poorer children in rural areas are less likely to go to school	Poorer children are not motivated to participate in formal education (because they or their parents have had negative experiences at school).	Yes. Policies will provide online learning programmes, and digital skills training.	Yes. Measuring the number of schools connected to the internet and implementing the nationwide roll out of a digital skills curriculum.	No. Digital content and skills training does not necessarily lead to increased motivation to participate in formal education; the KPIs do not measure participation in formal education.

Note: The two scenarios presented are already very advanced; often policies do not identify the causes underpinning the problem and do not identify clear measurable KPIs.

Figure 5 provides an example of how this might be visualised for a policy that aims to tackle the problem of children in rural areas being less likely to go to school.

b) Analyse representations of the group of interest

In addition to the coding using distinct categories for stakeholders, beneficiaries, delivery mechanisms, outcomes and mentions of outcomes (e.g., child rights), a more in-depth analysis of representations of the group of interest should be done.

Techniques here are akin to **qualitative discourse analysis**, looking at the overall positioning of children vis-à-vis adults and other (vulnerable) groups in society. This part of the analysis delves deeper into the question of **whether policies engage with the group of interest in a meaningful way**, allocating agency to them, with not only rights to be protected by others but also as individuals or a group with power to make and be involved in decisions about issues that impact them directly.

The analysis for the DFC project initially assumed that there were ways in which children were represented in policies: as **digital natives**, **victims** or **agents.** In-depth analysis suggested that the representations of children as digital natives and children as agents were not that useful to classify digital policies. After discussions between researchers, it was decided to replace these with analyses of discourses around children as digital resources (with restricted agency and without decision-making power) and children as rights holders (with more agency and power).

Final analyses of representations of children across policy documents was thus guided by three types of representations (in order of prevalence): children as **resources**, children as **victims** and children as **rights holders**.

Based on the thematic coding presented in the previous section, these were linked to the goals as conceptualised in the SPG framework (i.e., economic development, social inclusion, civic participation and individual rights) and the associated beneficiaries, digital delivery mechanisms, outcomes and stakeholders.[18]

In the analysis of inequalities, you examine in more detail who the beneficiaries of policies are said to be and whether your group of interest (e.g., children) are seen as all the same or whether there are inequalities between them (e.g., gender, disability, social class, caste, ethnicity, tribe, in formal education or employment).

[18] Helsper et al. (2025a); Liu et al. (2024).

After examining how children were seen as a group, the DFC project looked at whether children were seen as a **homogenous group** or **if inequalities** between children from different backgrounds (the analogue environment) were considered.

For the DFC project it was important to also analyse whether any explicit or implicit reference was made to children's ages or developmental stages (e.g., infants, minors, babies, teens, students, youth) and age-appropriate design. You should look at whether there is an **intersectional understanding of disadvantage**, for example, whether all girls are supposed to be the same or whether there is a difference depending on where they live (e.g., in rural areas) or their citizenship status (e.g., migrants). In your analysis this should be linked to policy discourses, ideologies and power dynamics that might underpin policy construction or design.

All the codes, qualitative and descriptive analyses described in this toolkit come together for you to be able to draw conclusions about which discourses or narratives underpin policies related to your group of interest. This process of discourse analysis is iterative, and discourses are only likely to emerge after repeated reading of policies and cross-policy comparisons. In this part of the analysis in particular, discussion between researchers is needed, and knowledge of national, country or local context is fundamental. Detailed qualitative note keeping and collection of longer quotes is key to analysis of this kind.

Supplementing your analysis with other methodologies

The methodology presented provides a wide range of insights, allowing the designer, researcher or evaluator of policies to **identify and address strength and weaknesses of digital policies** and the **representation of different groups**. It is also flexible and adaptable to different context and can be easily supplemented with other qualitative or quantitative analytical techniques. Further in-depth **qualitative analytical techniques**, such as **critical discourse analysis**,[19] can provide the historical and current political and ideological context for selected country or regional case studies.

Another potentially useful technique is **cosine similarity** mapping, which allows for analysing discourse patterns in large policy corpuses. For the DFC project, this type of analysis was tried out for digital policies from China, the report on which can be found as an **addendum** to this toolkit. As a computational linguistics method, similarity mapping applies word-embedding models to policy discourse analysis. It shows how policy language groups certain ideas together by examining whether different terms tend to appear in similar linguistic contexts across many policy documents in the corpus. This makes the semantic similarity between words observable as measurable semantic proximity.

[19] Burchell et al. (1991); van Dijk (1985).

Cosine similarity mapping and other **quantitative linguistic techniques** help reveal how certain policy subjects are represented in a large corpus of policy texts: for example, whether they are placed closer to concepts such as talent development, citizenship or safety. It complements close reading of key policy documents by revealing broader semantic structures that are not visible through qualitative interpretation alone.

Step 4: Reporting and dissemination

This methodological toolkit provides a comprehensive framework for analysing digital policies and how they represent groups of interest to the researcher. Its strength lies in its adaptability across different contexts – not only in its application, but also in the types of outputs it can produce to generate impact.

At this last key step, it is important to tailor your output to the stakeholder and make clear what the key 3-7 messages are that should be of interest to them. You need to identify not only which stakeholders should be interested, but also what their priorities are, and what language and types of outputs they are likely to use and engage with. We finish this toolkit by identifying three potential audience groups that may benefit from your research, with strategies for tailoring your outputs for impact. We combine steps 4a (audience) and 4b (impact) from Figure 1 in this section.

1. Policymakers and government officials

For policymakers, you should do the following in your outputs:

- Keep it short and to the point, and do not expect these stakeholders to read beyond the executive summary and recommendations.
- Emphasise clear action points, with outputs in the form of research reports with executive summaries that highlight critical gaps, priority areas for intervention and specific legislative or regulatory changes needed.
- Frame recommendations in terms of best practice, problems in compliance with international or local standards and alignment with existing policy priorities.
- Select key quotes that make your findings come to life; presentations of statistics in terms of how many times certain groups or concepts are mentioned and the (in)adequacy of KPIs are likely to be noticed.
- (If available) Provide evidence of public support for policy changes or international best practices, strengthening the potential for impact.

2. Civil society organisations and advocacy groups

For civil society, specific recommendations and resources for advocacy and campaigning opportunities are useful and will increase impact. In publications you should:

- Highlight contradictions between stated commitments to broader social goals and actual policy design and implementation.
- Provide evidence-based talking points; using quotes and highlighting points that are media-friendly (i.e., aligned with recurrent myths, current policy discussions or hot topics).
- Identify opportunities for coalition building by pointing out other organisations could be involved.
- Provide supplementary outputs such as practical tools, such as frameworks for effective engagement with public consultations.
- Design flyers or short animations with key findings that organisations can use for their volunteers, professionals and beneficiaries.

3. Academic and research communities

Researchers are likely to read the grey literature, but will look for academic publications that are peer-reviewed and available under open access copyright schemes. You should include the following in your publications for an academic audience:

- Provide extensive argumentation around the theoretical underpinning your research.
- Report in detail on your methodological challenges and learnings, and future research directions as identified from your work, will add to the emerging body of work in this field.
- Write conclusions detailing what is new about what you found and what further research is needed.
- Include in outputs and presentations of your work how academics might impact policymakers or work with other non-academic stakeholders so that their work has real-life impacts.

Concluding remarks and request for feedback

This completes all the steps in the methodological toolkit. They will take you from corpus construction to communicating your findings to a broader public.

The methodology put forward by this toolkit is designed to evolve with experience and feedback, and researchers are encouraged to document their adaptations and share insights with the Digital Futures for Children centre (DFC).

Please let us know if you find this useful, have applied it in your own work and whether there is anything that can be improved in the next addition of this toolkit.

If you have any further questions or comments, contact Ellen Helsper at the London School of Economics and Political Science, or the DFC: www.digital-futures-for-children.net/contact-us

Addendum: Similarity mapping of digital policy discourses – Case study China

Anran Wang

This section presents a computational linguistics method for analysing how children are represented in national digital policy discourse. Using Chinese central government digital policies (2020-24) as a case study, the study demonstrates how word-embedding models can be used to

- construct a text-based representation of policy discourse;
- translate conceptual imaginaries into measurable linguistic anchors; and
- generate interpretable indicators such as cosine similarity scores between child-related terms and key policy concepts.

In the empirical illustration, the analysis identifies differentiated semantic patterns for the three terms most commonly used in these policies to denote children (儿童– children, 学生– students and 未成年人– minors), showing how semantic proximity reveals distinct policy discourses attached to each reference. In other words, the same population – children under 18 – is discursively partitioned into separate categories, each with a different policy logic.

Background

Digital policies assign meaning to children through the concepts that surround them: safety, skills, participation, rights and responsibility. These meanings are rarely formal definitions. They emerge through repeated linguistic associations distributed across diverse policy documents issued for different regulatory and developmental purposes. Understanding how imaginaries of children are formed, therefore, requires attention not only to what individual policies say, but also to the patterns that link them as policy discourse.

Here we examine these patterns using a corpus-based computational linguistics approach. By modelling the semantic proximity between 'children' and other policy themes, word-

embedding modelling makes it possible to describe how representations are organised in the language of digital policy. Instead of interpreting representations only as narratives or codes (themes), the approach treats discourses as measurable relations within a corpus: which ideas appear closest to children, which domains they cluster with and how strongly these associations are expressed across documents. The result is an interpretable set of indicators that translates discursive patterns into reproducible analytic outputs.

A corpus of Chinese central government digital policies (2020-24) is used as a case study. These policies are available in a publicly accessible repository, presented in full text and formatted in a consistent way, which makes it feasible to compile them into a searchable corpus and apply word-embedding modelling without extensive preprocessing. The workflow itself is not specific to this corpus; where similar repositories and tokenisation tools (for breaking text into analysable units) exist, the same method can be applied to examine how representations of a group of interest, in this case, children and child rights, are organised within different policy contexts.

Corpus construction

The corpus consists of central government digital policy documents in China between 2020 and 2024, a time window that aligns with the broader project's comparative scope. All documents were sourced from the State Council Policy Document Library,[20] which archives full-text policies issued by the State Council of China and its affiliated national ministries. At the time of analysis, only State Council–level documents could be reliably collected in full text, so the present dataset is limited to this level of policymaking.

Policies drafted between 1 January 2020 and 31 December 2024 were retained if their full text contained at least one of the following terms: 'internet' (互联网), 'digitisation' (数字化), 'informatisation' (信息化), 'artificial intelligence' (人工智能) or 'network' (网络). Keyword matching was applied to the full text rather than to policy titles, as word-embedding models require a reasonably sized corpus to produce stable semantic representations.

Out of 611 policy documents issued during this period, 379 met these criteria: 80 from 2020, 91 from 2021, 64 from 2022, 60 from 2023 and 84 from 2024. Although these numbers fluctuate year by year, they indicate a relatively steady level of digital policy activity across the five-year window. The corpus covers major policy domains, including infrastructure, governance, public services, industry, education and cybersecurity, providing a sufficiently broad basis for examining how digital policy imaginaries are articulated across different sectors of the Chinese policy system.

[20] Chinese State Council Policy Document Library: www.gov.cn/zhengce/zhengcewenjianku

Text processing and model training

Because Chinese does not use spaces to mark word boundaries, the texts first required word segmentation to split continuous characters into meaningful units prior to modelling. The full text of each document was then cleaned using a custom written script to remove punctuation, numbers, non-Chinese characters and common stop-words. Single-character tokens were also dropped, as they rarely carry distinctive semantic value in policy settings. The resulting corpus provides a coherent linguistic dataset suitable for training domain-specific word-embedding models. In total, the processed corpus contains 906,058 words and 25,950 unique word types, yielding a vocabulary density of 0.0286, meaning that around 2.9% of all tokens are unique word types, which suggests adequate lexical diversity for reliable semantic modelling.

Word2Vec is a widely used method for generating word embeddings—numerical representations of words based on how they are used in context. A skip-gram Word2Vec model[21] was then trained on the corpus to generate semantic representations of policy terms[22]. The skip-gram architecture is well suited to small, domain-specific corpora, and effectively captures contextual relations among words. Hyperparameters widely used for training Word2Vec models on small- or medium-sized corpora were also applied here: a 200-dimensional vector space, a context window of five words, and a minimum frequency threshold of five.[23] Those wishing to replicate the modelling should use these hyperparameters.

To confirm model validity, pilot tests examined similarity among conceptually related terms within the policy lexicon (e.g., ‘education’ [教育], ‘school’ [学校], ‘student’ [学生]). The model placed these terms near each other, indicating that it captures meaningful policy-related relationships rather than random co-occurrence.

Lexical patterns of references to children

Before examining semantic relationships in the embedding space, it is useful to consider how children are referred to in Chinese policy language. Following the UNCRC definition of children as persons under 18 years of age, the analysis adopts a broad operationalisation that includes terms denoting childhood, schooling and legal minority status.

[21] Word2Vec commonly uses two architectures: Continuous Bag of Words (CBOW), which predicts a target word from its surrounding words, and skip-gram, which predicts surrounding words from a target word. Skip-gram is generally preferred for smaller, domain-specific corpora because it learns stable semantic relationships even when word frequencies are uneven.
[22] The Word2Vec analyses was conducted on the Gensim library in Python.
[23] Mikolov et al. (2013).

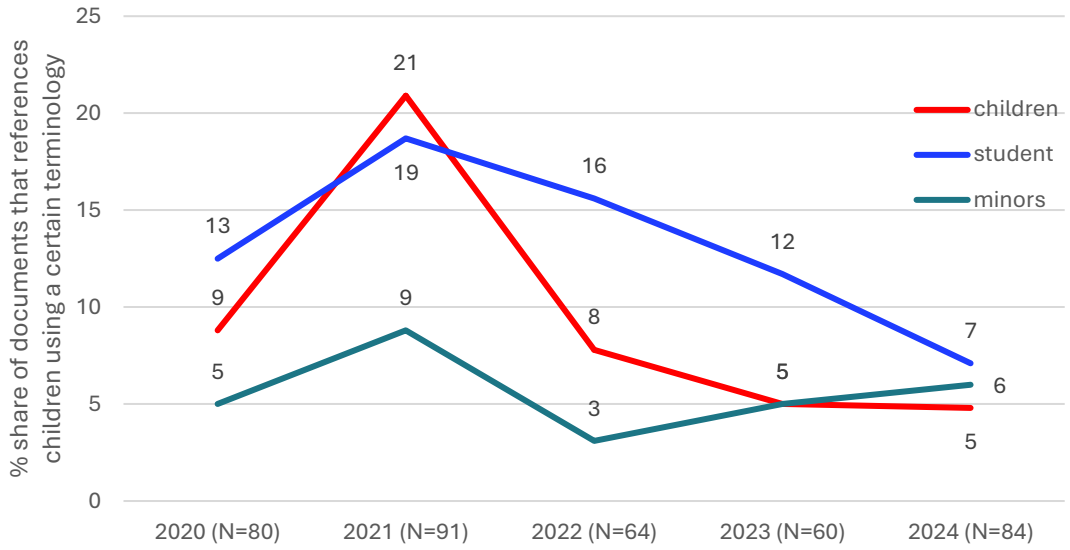
Across the 379 policy documents in the corpus, three expressions – 'children' (儿童), 'students' (学生)[24] and 'minors' (未成年人) – account for the vast majority of explicit references to children. Their frequencies are summarised in Table A1.

Table A1: Frequency of terms referring to children in 379 central government digital policy documents (2020-24)

Terms	Total occurrences	Documents mentioning the word	Share of documents (%)
'Children' (儿童)	458	40	10.6
'Students' (学生)	276	49	12.9
'Minors' (未成年人)	240	21	5.5

The overall frequency of these terms is modest relative to the full corpus. 'Children' accounts for the largest number of total occurrences, while 'students' appears in the greatest number of documents. 'Minors' is comparatively rare on both measures.

Figure A1: Mentions of children, students and minors in Chinese digital policy documents (share of documents)



Given that policy texts tend to use these references in different ways – for example, 'children' appears more often in wellbeing- and family-related discussions, 'students' in education and training and 'minors' in regulatory language – these observations suggest that the policy discourse around children is not linguistically uniform across policy

[24] In Chinese policy documents, "学生" can technically include university students, but the term "大学生" is normally used when higher-education students are meant. Unqualified references to "学生" usually refer to children in primary and secondary schooling.

domains. These representations also vary over time (see Figure A1). Mentions of children in all categories have decreased, with a peak in the pandemic year 2021.

The understanding and use of these terminologies provides an empirical baseline for the similarity mapping analysis in the next section.

Similarity mapping of representational discourses of children

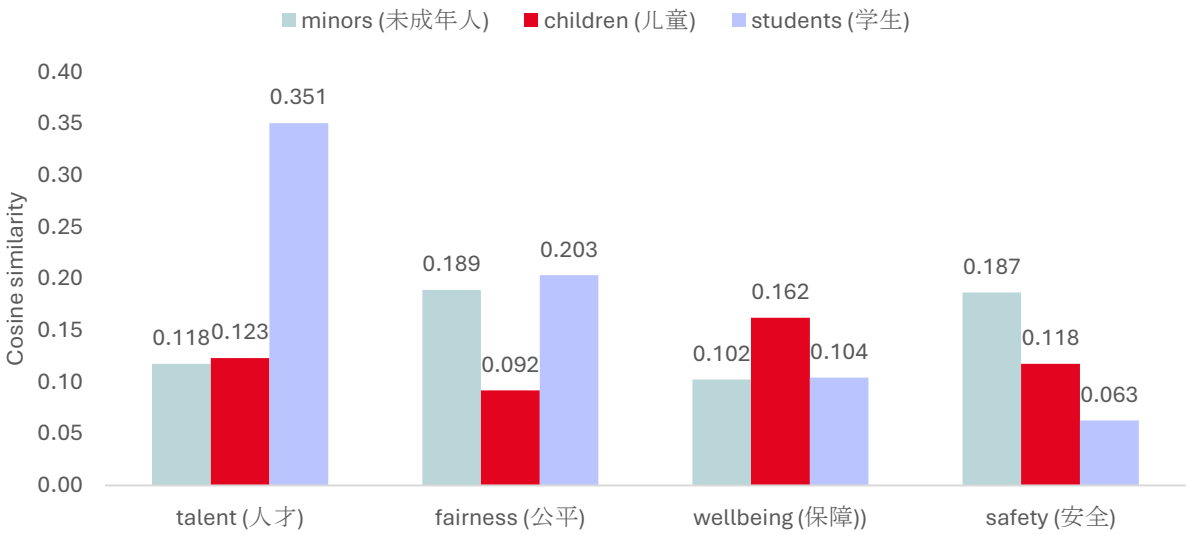
Similarity mapping uses the trained word-embedding model to examine how references to children are positioned relative to key policy themes. Instead of counting co-occurrences within individual documents, this approach estimates how frequently two terms appear in similar linguistic contexts across the entire corpus. From these patterns, cosine similarity scores are derived to capture the relative proximity between words in the policy corpus.[25]

To operationalise these policy themes in measurable linguistic form, four anchor terms were selected: 'talent' (人才)[26], 'fairness' (公平), 'wellbeing' (保障)[27] and 'safety' (安全). Their selection reflects both the Social Policy Goal (SPG) model's distinction between economic, social inclusion and rights discourses and the actual linguistic usage in the Chinese policy corpus.[28] Notably, these anchors do not map neatly on to English academic conceptual categories; moreover, some dimensions – for example, 'civic participation' – are scarcely expressed with children as a grammatical subject in this corpus (similar to what was found in the global review of policies). The selected anchors emerged from pilot tests as the most robust operational choices to distinguish different meanings, although they remain imperfect linguistic approximations of the broader representations of children and children's rights.

The similarity analysis compares the three reference terms – 'children', 'students' and 'minors' – against the four policy anchors: 'talent', 'fairness', 'wellbeing' and 'safety'. The resulting cosine similarity values (see Figure A2) provide a comparative view of how each reference is positioned within China's digital policy discourse, highlighting which policy imaginaries are most closely associated with each term. Values close to 1 indicate strong association, values near 0 suggest weak or no association and negative values imply contrasting contexts.

[25] Jurafsky et al. (2025).
[26] The choice of "talent" reflects the common use of "人才" in Chinese digital-policy discourse, where children are imagined as future contributors to innovation and national economic development. This mirrors the SPG framework's economic-development dimension (see Helsper et al. 2025a).
[27] The term 保障 is widely used to state the government's responsibility to ensure that certain conditions or services are in place. Its literal meaning is closer to "safeguard", but its actual usage is broader, not limited to state benefits and related to a broader understanding of care. Here, "wellbeing" is used as the closest English equivalent.
[28] Helsper et al. (2025a).

Figure A2: Cosine similarity between child-related terms and policy anchors



The similarity results reveal distinct semantic configurations across the three reference terms. 'Children' is most closely associated with 'wellbeing' (0.162), suggesting a policy imaginary that foregrounds care and family support. 'Students' displays a markedly different profile, with its strongest similarity to 'talent' (0.351) and 'fairness' (0.203), reflecting an emphasis on education, skills development and equal access to learning opportunities. 'Minors' is most strongly associated with 'safety' (0.187) and 'fairness' (0.189), indicating a more regulatory framing. While these patterns may not be surprising, they underscore an important linguistic feature of Chinese policy discourse: the same population – children under 18 – is described through multiple reference terms, each carrying a different policy imaginary.

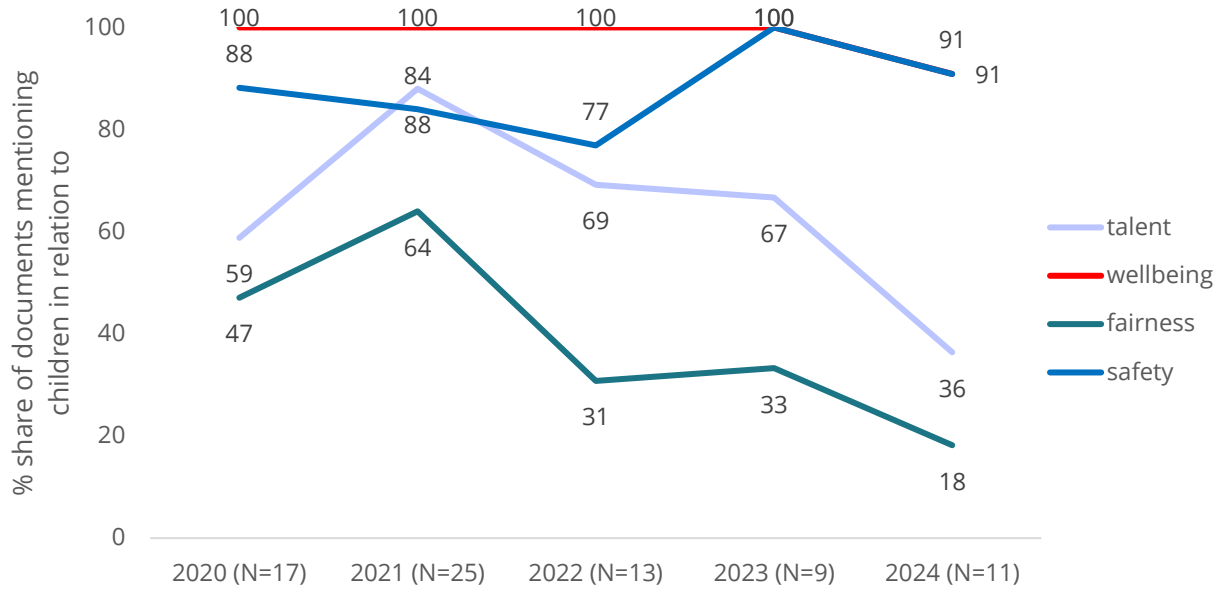
Table A2: Mentions of children in relation to key discourse terms over time

Type of mention	Associated term	2020	2021	2022	2023	2024
Total number of documents in which children are mentioned		17	25	13	9	11
Documents mentioning children in relation to:	Talent	10	22	9	6	4
	Wellbeing	17	25	13	9	10
	Fairness	8	16	4	3	2
	Safety	15	21	10	9	10
Total number of times children are mentioned in relation to:	Talent	98	269	105	96	9
	Wellbeing	295	651	238	112	79
	Fairness	17	73	7	6	2
	Safety	112	684	105	93	173

Further analysis showed that the general discourses in policies that mention children vary over time. Children were mentioned in fewer policy documents in 2023 and 2024 than in 2020 and 2021 (see Table A2).

While children consistently appear in relation to policy texts on wellbeing, with a small drop in 2024, other associations are more variant over time (see Figure A3).

Figure A3: Representation of children in different contexts over time (share of documents)



Discussion and limitations

The similarity mapping analysis helps contextualise the patterns reported in the China section of the report *Left out and misunderstood: Children in global, regional and national digital policies*,[29] which notes that Chinese digital policy tends to frame children either as resources for future economic development or as vulnerable users who require protection. The similarity mapping is consistent with this observation, but adds an important nuance: the policy imaginary depends on which 'type of child' is referred to in the policy.

Although 'students' and 'minors' dominate in flagship digital policy documents, the term 'children' is actually the most frequently used reference across the full corpus of 379 documents. When 'children' is explicitly used to denote the population, it is most strongly associated with wellbeing – suggesting a representation of children as recipients of services and support rather than as economic resources or victims. This implies a non-agentic form

[29] Helsper et al. (2025b).

of citizenship, perhaps more aligned with how citizenship is understood within the Chinese context.

The divergence may further be shaped by the structure of policymaking in China. Under the State Council, multiple ministries have formal responsibilities for children's rights and wellbeing, and they issue many policy documents where digital technologies are framed as a means to protect and support children's rights. Yet these ministries do not appear to be central actors in China's digital policy planning. As a result, 'children' are highly visible in State Council wellbeing-oriented documents, but only partially visible in core digital ones. The fragmentation of reference terms – 'children', 'students', 'minors' – therefore mirrors a fragmentation of institutional responsibility, helping to explain why some children and children's rights can be left out or misunderstood in China's digital policy discourse.

Part of this divergence also arises from methodological constraints: because word-embedding modelling requires a sufficiently large corpus, the dataset includes many digitally relevant policies that are not flagship digital policies on which the analysis was based for the global review report.

References

Burchell, G., Gordon, C., & Miller, P. (eds.) (1991). *The Foucault effect: Studies in governmentality.* University of Chicago Press.

Helsper, E. J. (2021) *The digital disconnect: The social causes and consequences of digital exclusion*. SAGE Publications.

Helsper, E. J., Rao, S., & Lyons Longworth, M. (2025a). *Left out and misunderstood: Children in digital policies. A global review.* Digital Futures for Children centre, LSE and 5Rights Foundation. https://eprints.lse.ac.uk/130444/

Helsper, E. J., Rao, S., & Lyons Longworth, M. (2025b). *Left out and misunderstood: Children in global, regional and national digital policies.* Digital Futures for Children centre, LSE and 5Rights Foundation. https://eprints.lse.ac.uk/130445/

Jurafsky, D., & Martin, J. H. (2025). *Speech and language processing (3rd ed.).* Stanford University Press. https://web.stanford.edu/~jurafsky/slp3

Liu, S., Wang, A., & Helsper, E. J. (2024). *Four models of digital inclusion policy making.* Paper presented at the 74th International Communication Association Framework Conference, Gold Coast, Australia.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space.* arXiv:1301.3781. https://doi.org/10.48550/arXiv.1301.3781

van Dijk, T. A. (Ed.). (1985). *Discourse and communication : New approaches to the analysis of mass media discourse and communication.* de Gruyter.

Digital Futures For Children

Research at LSE and 5Rights Foundation

digital-futures-for-children.net
info@dfc-centre.net

@5RightsFound @MediaLSE
#DFC #DigitalFutures4Children

The Digital Futures for Children centre acknowledges funding from the 5Rights Foundation.

This joint LSE and 5Rights research centre supports an evidence base for advocacy, facilitates dialogue between academics and policymakers and amplifies children's voices, following the UN Committee on the Rights of the Child's General comment No. 25.

Please cite this research report as: Helsper, E. J., Rao, S., & Lyons Longworth, M. (2025). *Digital policy analysis: A methodological toolkit.* Futures for Children centre, LSE and 5Rights Foundation. https://eprints.lse.ac.uk/130446/

Cover image: Freepik