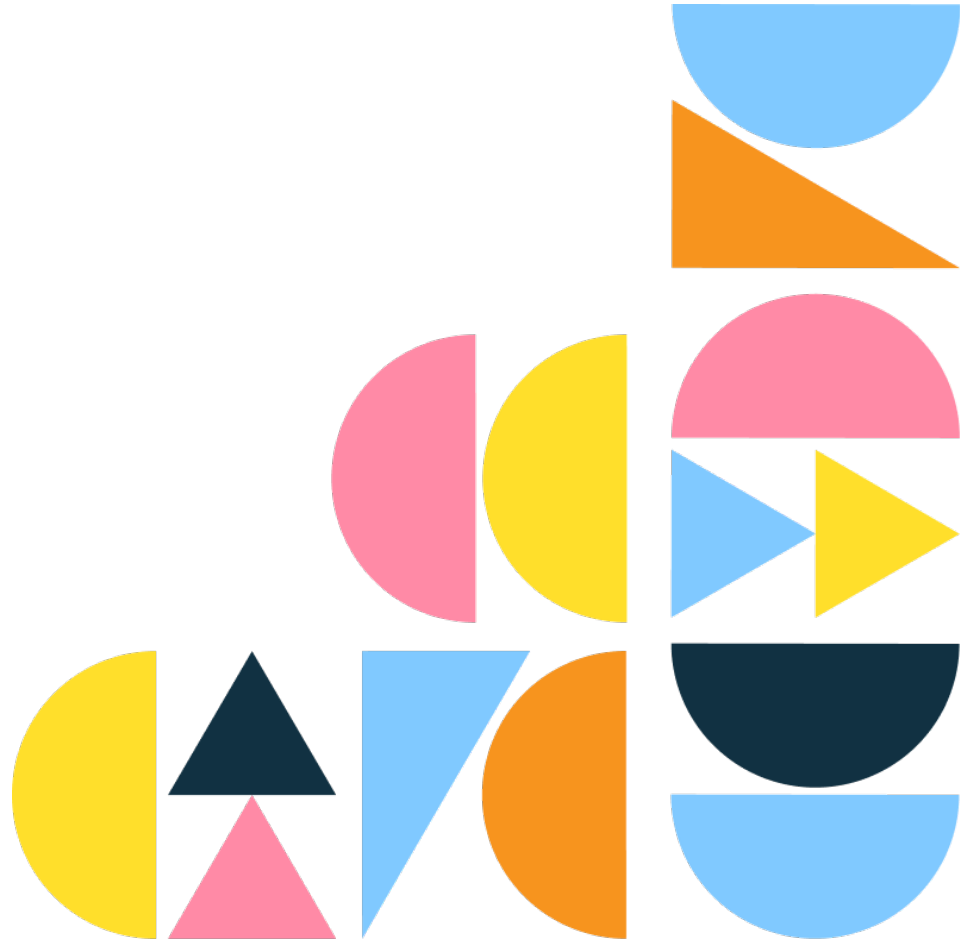


# Prevalence of Potentially Prohibited Items on Search Services

Final Report

*18 September 2023*

**PUBLIC**



**CONTENT WARNING**

*This report and accompanying materials discuss the sale and supply of potentially prohibited items, such as Bladed and Non-bladed Offensive Weapons, Controlled Drugs and Psychoactive Substances as defined under current United Kingdom legislation. This is strictly for the purposes of enabling research. Please be aware that the content and activity discussed could be offensive and may cause you harm.*

# Table of Contents

## Sections

## Reference

---

### Introduction and Context

*Context of the project and introduction to research questions and controlled variables*

Pages 6 - 11

---

### Methodology and Scoring

*Methodological approach and limitations*

Pages 12 - 25

---

### Summary of Findings

*Top level insights answering research questions on prevalence of potentially prohibited items, across search services, potentially prohibited item categories, search service products and features*

Pages 26 - 36

---

### Further Insights

*Extended analysis, including qualitative insights, categorical variables and findings from indicative analysis of the prevalence of potentially prohibited items on DuckDuckGo*

Pages 37 - 47

---

### Functional Analysis

*Breakdown of search service user journey by functionality, examining features such as Autocomplete, Related searches and Sponsored results, as well as different Results page products*

Pages 48 - 59

---

### Reflections, Implications

*Challenges and implications of the research*

Pages 60 - 62

---

### Appendix

- *Further details methodology*
- *Further details on trends*

Pages 63 - 69

# Executive Summary – what we did (1)

- In March 2023, Ofcom commissioned PUBLIC to deliver a project to build a robust evidence base regarding the **prevalence and ease of access to content** that contains an apparent offer to sell or supply a range of potentially **prohibited items or articles**, accessible via **online search services**. This is referred to as **potentially prohibited content** throughout this report. This report summarises the key findings from the research project.
- A total of **384 searches were conducted**, focusing on a range of potentially prohibited items. Each unique search query was tested on the two major search services in scope of the research, **Google** and **Bing**. For each individual search, the research sought to identify how many webpages within a limited sample could be classified as meeting the criteria for containing potentially prohibited content.
- The queries incorporated **4 categories** of potentially prohibited content: 1) knives and bladed weapons, 2) firearms, 3) controlled drugs and 4) psychoactive substances. **4 specific items** were selected per category. Each query was searched across **4 search service products**: text search, image search, shopping, and video search.
- For each of the 384 searches, **12 result URLs** were selected at random from the **top 30 results** and analysed, which led to **4,608 results**.
- A **three-part scoring system** was developed to classify webpages against criteria for containing potentially prohibited content. The criteria was used to assess whether webpages contained the following:
  - The presence of potentially prohibited item as a product
  - The presence of a clear route to purchase
  - Accessibility to someone resident in the United Kingdom.

## Executive Summary – what we found (2)

- Our research found that **potentially prohibited content was often present within webpages directly accessible from search engine results pages.**
  - The research **found webpages that were classified as containing potentially prohibited content** within the sample of URLs reviewed for every potentially prohibited item tested, **across both search services.**
  - Across all research categories, searches for **Psychoactive Substances** returned the most webpages meeting the criteria for containing potentially prohibited content, while searches for **Firearms** had the least.
- **Autocomplete, search engine results pages, search products, and related and sponsored searches** were all identified as **functionalities that play a role** in surfacing the potentially prohibited content researchers identified in the sample.
  - The research **found webpages containing potentially prohibited content in sponsored results for some of the search queries tested.**
  - Both search services presented researchers with **relevant ‘autocomplete’ and ‘related searches’ suggestions** when conducting searches for potentially prohibited items
- Of the four search service products tested, **text search returned the highest number of webpages classified as containing potentially prohibited content.**
  - Overall, **video** and **shopping search service products** returned **comparatively lower numbers** of these kind of webpages.
- Using an **extended search query** – one that was more detailed and specific – **returned the greatest number of websites classified as containing potentially prohibited content**, compared with more general types of search query tested.
  - This was likely due to the use of **targeted slang** and **coded language** in the query, as well as the inclusion of **additional purchasing-related language.**
- Webpages meeting the criteria to be classified as potentially prohibited content tended to **appear higher up in the search engine results page** – with more URLs classified in this way coming from search results 1-10, compared to either 11-20 or 21-30.

# Introduction and Context

*Potentially Prohibited Content and Search Services*




*(Pages 6-11)*

The Online Safety Bill, once in force, will place duties on search services to assess the risk of users encountering offers of sale or supply of potentially prohibited items.

This section provides the context and background to this project and its objectives. We outline:

- **Regulatory and sector contexts** informing this project
- The **Research Questions**, objectives and controlled variables chosen in order to investigate them
- **Legal, policy and technical caveats** to the methodology and findings of the project

# We performed an initial trend analysis to understand the body of existing research and source preliminary themes

Theme	 <b>Search service market</b>	 <b>Online harm accessibility on search services</b>	 <b>Potentially prohibited content</b>
Trend	<p>1. <b>Google has a very large share of the UK search service market<sup>1</sup></b>, while Bing holds significant share and <b>others command less than 1% of the market.</b></p> <p>2. The functionality of search services has <b>evolved to include multiple widgets</b> on the search engine results page<sup>2</sup> but there is <b>limited research into the implication of this</b> on the availability of potentially prohibited content.</p>	<p>3. <b>Minimal research<sup>3,4</sup> has been conducted on the prevalence of potentially prohibited content on search engines</b>, with adjacent research interrogating suicide and self harm<sup>5</sup> or child sexual abuse material.</p>	<p>4. Research has found that potentially prohibited content is more likely to be found from (i) <b>direct messaging</b> or dedicated social media groups<sup>6</sup>, (ii) the <b>encrypted deep web</b> (e.g. the Onion Router), or (iii) <b>cryptomarket forums<sup>7</sup></b> (e.g. Pretty Good Privacy).</p> <p>5. There is <b>very limited evidence and research</b> into the accessibility and trading of <b>firearms<sup>8</sup> or bladed weapons</b>. This is likely <b>as a result of the research not being conducted.</b></p> <p>6. There is a greater volume of research into the provision of <b>controlled drugs and psychoactive substances<sup>9</sup></b>, identifying a wide variety of evasive techniques used by perpetrators.</p>

For extended trend analysis, see **Appendix**




Source(s): (1) [Similarweb](#); (2) [From 10 Blue Links Pages to Feature-Full Search Engine Results Pages](#); (3) [Online Content Study](#); (4) [Do Search Engines Influence Media Piracy? Evidence from a Randomized Field Study](#); (5) [How Search Engines Handle Suicide Queries](#); (6) [Sky News](#); (7) [An evidence synthesis of strategies, enablers and barriers for keeping secrets online regarding the procurement and supply of illicit drugs](#); (8) [Criminal armouers and illegal firearm supply in England and Wales.](#); (9) [Internet-facilitated drugs trade](#)

# This project aimed to build an evidence base on the risk of potentially prohibited items being accessible on search services






## Challenge

This project aims to build a robust evidence base regarding the existence, prevalence and ease of access to content that contains an apparent offer to sell or supply a range of potentially prohibited items or articles, accessible via online search services.

### KEY

-  Quantitative Methods
-  Qualitative Methods
-  Pattern Analysis

## Research Questions

-  Is content related to the sale or supply of potentially prohibited content accessible 'one click' away from the search service results page?
-  What is the prevalence of potentially prohibited content within search results delivered by search services?
-  Does functionality on search services play a role in surfacing potentially prohibited content, and if so, how?
-  How, if at all, do answers to the above differ between search services, and their products?
-  What, if any, patterns are there in the potentially prohibited content that is accessible from search services?



# The definitions for potentially prohibited content and 'one click' are important framing for interpretation of results

## Potentially Prohibited Content (PPC)

For the purpose of this research, potentially prohibited content is content featuring “potentially prohibited items” (in this case, weapons, firearms, illegal drugs or psychoactive substances – as referenced in Schedule 7 of the Online Safety Bill in relation to certain “Priority Offences”) where certain indicative factors (as set out in slide 20) are present to suggest that the items are being marketed for sale to the UK.

## 'One click'

For the purposes of this research, being 'one click' away from the first search engine results page relates to the webpage that can be found directly from opening the link provided by a search engine following a query. Only the landing page that opens when a link is followed was analysed for this research. For the avoidance of doubt, this means that no other pages – even if on the same website – were assessed against scoring criteria.

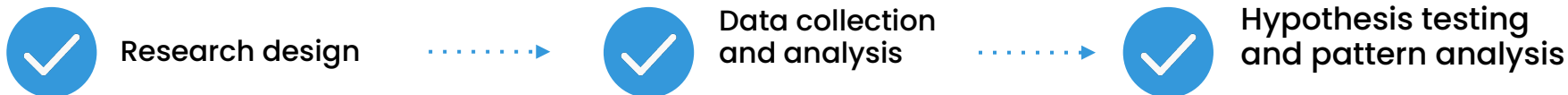
## "Sale"

For the purposes of this research, reference to the “sale” of PPC encompasses a wide range of conduct, including: offering to supply, advertising, marketing, and making available for sale. These definitions and the indicative factors used in this research have been created and applied for the specific purposes of this research exercise.

## Prevalence

This report considers “prevalence” at various stages. For the avoidance of doubt, prevalence in this report refers to the frequency of PPC being identified as it is defined by the preceding definitions.

# PUBLIC's three-step plan was designed to deliver meaningful answers to the project's five key research questions



In the first weeks of the project, PUBLIC established the basis of the research by:

1. **Gathering evidence** on search services, potentially prohibited content and past research into search services and potentially prohibited items, including methodologies
2. **Designing a research approach** that covered all of Ofcom's research questions efficiently
3. **Iterating this research design** using a workshop with Ofcom to ensure we are aligned with priorities

We assessed the prevalence and presence of potentially prohibited content by:

1. Establishing **keywords and search queries** for potentially prohibited content categories
2. **Executing** these across major search engines, across prioritised products
3. **Recording, annotating and scoring the results** by agreed guidelines and metrics

We sourced findings by:

1. Analysing the **effects of search engine results page functionality**
2. **Identifying quantitative and qualitative patterns** across search services and potentially prohibited content categories using scores
3. Using the functional analysis of different result pages to **identify the impact of design**
4. **Conducting** additional agreed upon specific **hypotheses to supplement findings**

# Despite a robust project plan, we should bear in mind the limitations of the study

## Policy

During the research there were **updates in policy** that may have affected findings across some categories of potentially prohibited items.

In April 2023, the Home Office launched a [consultation on new knife legislation proposals to tackle the use of machetes and other bladed articles in crime](#). This was **announced during the research period**, and due to the high volume of relevant news articles released at the time, may affect findings to do with Bladed weapons.

## Technical

You can refer to a detailed breakdown on the limitations of our methodology at slide 25. Some of the points include:

- **Small sample sizes** mean that comparisons should not be made at the smallest unit of analysis (i.e. multiple variables compared at one time)
- Results from this research should be cited with **appropriate context**.
- **Scoring** is a subjective indicator of likelihood, rather than evidence of an offence.

## Legal

"Illegal content" is a new legal concept created by the Online Safety Bill and refers to content that amounts to a "relevant offence", as defined under the Bill (clause 59). There are "relevant offences" relating to the sale of weapons, firearms, illegal drugs or psychoactive substances – as set out in Schedule 7 of the Bill under "Priority Offences".

For some of these items, it is fairly straightforward to determine whether the online marketing of them is potentially illegal content. For others, it is less clear, because whether content could be considered 'illegal' will depend on offline circumstances too. For the purposes of this research, we have looked for content featuring these items where certain indicative factors (as set out in slide 20) are present to suggest that the items were being marketed for sale to the UK. We refer to this content as "potentially prohibited content", and the items featured in the content as "potentially prohibited items".

# Research Project Overview & Methodology

(Pages 12-25)

This section covers the research methodology step-by-step, through 6 steps across two phases:

## Phase 1

*Research design*

1.1 Evidence building and Taxonomy

1.2 Research Design

1.3 Trial and Validation

## Phase 2

*Field work*

2.1 Data Collection

2.2 Data Analysis and Quality Assurance

2.3 Additional Analyses

# The methodology controlled specific variables to answer the research questions with a sample of ~4,600 URLs

## Variable factors

We chose specific factors to vary in order to explore the research questions.

**RQ**      **Variable(s)**

1 2  
5

Potentially prohibited item categories and potentially prohibited items



1 2  
3 4

Search Service Products



1 2

Search Services



1 2  
5

Search Queries



## Controlling Variables

We consulted with Ofcom on how these factors would be varied, e.g. Number of search services to explore

### 4 potentially prohibited item categories

*Prohibited knives/bladed weapons; Prohibited offensive weapons; Controlled drugs; and Psychoactive substances*

### 4 potentially prohibited items per category

### 4 Search Service products

*Text, Image, Shopping, Video*

### 2 Search Services

*Google, Bing*

### 3 Search Queries per potentially prohibited item

## Results

How these variables translate to a dataset

For each of the 384 queries, **we analysed 12 result URLs** to establish whether potentially prohibited items were accessible



Total search results analysed:  
**4,608**

# We built a taxonomy of in-scope potentially prohibited items to establish a set of items to be researched, validated with Ofcom

## Longlisting of items

- **Four categories of potentially prohibited items** were selected by Ofcom for further research: Prohibited knives/bladed weapons; Prohibited offensive weapons; Controlled drugs; and Psychoactive substances.
- All items under these four categories were longlisted **using existing legislation<sup>1</sup> or official guidance<sup>2</sup>**.

## Taxonomy creation

Items were **grouped according to sub-categorisations** contained in the legislation or guidance they were sourced from to create a **taxonomy of potentially in-scope potentially prohibited items**. The full taxonomy can be found in the [Appendix](#).

## Shortlisting of items

- Items were shortlisted for possible research based on factors that ensured they would **provide a reasonable indication of prevalence** when searched for using search services.
- These factors included whether they were only **conditionally prohibited**, whether their **legal status was being reviewed**, or whether their **name was easily interchangeable** with non-potentially prohibited items.

## Ofcom validation

The shortlist of potentially prohibited items was **refined** in a workshop with **Ofcom subject matter experts and research specialists** to produce the list of four items in each category to be researched. These items can be seen in the [Summary of Findings](#) section of this report.

# Before executing the research, queries were tested to reflect likely real-world usage

To ensure the research reflects how users may search for items in the real world, each item was **queried in three ways**. All queries **began with “buy”** to reflect a **an attempt to sell/supply** the item and **ended with “uk”** to reflect the search being **geographically specific to the UK**, reflecting Ofcom’s geographic focus.

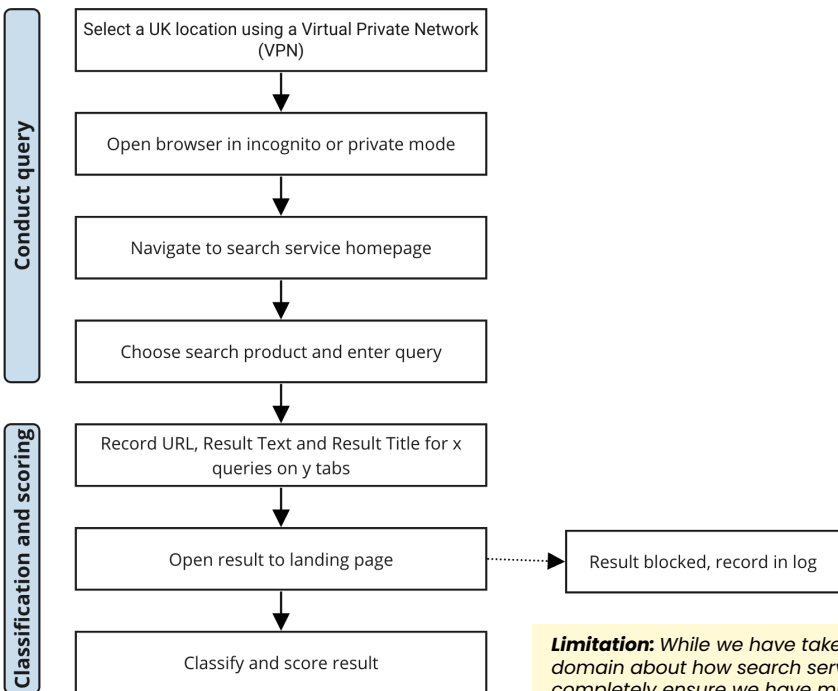
Query type	Description of language used in the query	Example query (Stun gun)
<b>Primary query</b>	Either the <b>term cited in the legislation or guidance</b> the item is prohibited in or, if this is not commonly used, the <b>most likely slang or coded term</b> <sup>1</sup> to be used to acquire the item as determined by the researcher	buy stun gun uk
<b>Secondary query</b>	The <b>slang or coded term most likely to be used</b> to find the item after the term used in the primary query.	buy taser uk
<b>Extended query</b>	Terms that used <b>more specific language</b> , such as a brand name or item descriptor, determined by the researcher to reflect how a knowledgeable user may attempt to acquire an item. Additionally, <b>“online” and “delivery” was included</b> to reflect attempts for the queried item to be purchased online and sent to an individual seeking to acquire them.	buy vipertek taser online uk delivery

# We designed a research process flow that factored in security risks and considerations for researchers

PUBLIC

1.2

## Process Flow



## Security Considerations

- A **Virtual Private Network (VPN)** was used by the researchers to ensure that searches were **not personalised** based on any prior activity or searching
- Searches were conducted in an **"incognito"** mode and collection of **cookies was opted-out** of to ensure **consistency of non-personalisation** across the research period.

## Takeaways

- This robust approach to security **enhanced the security of researchers** by mitigating exposure to malware and viruses.
- This process also **maintained the repeatability** of research during the project.

**Limitation:** While we have taken steps to mitigate risks of personalisation of results, there is limited information in the public domain about how search services track user behaviour and use this to inform results; therefore it is not possible to completely ensure we have mitigated all personalisation risks.



# We carried out a trial of the research methodology across a small sample of webpages for a final validation of approach

## Trial approach

PUBLIC used the week of 17 April 2023 to run a trial of the research parameters and process over ~250 webpages, to:

- Evaluate our chosen approach and validate our timing estimates
- Build an initial dataset, recognise preliminary trends and patterns in the results
- Propose any final alterations to methodology

Variables	No.
Search Services (e.g. Google)	1
Search Products	4
PPC Categories	4
PPC Item / Category	1
Queries / Item	3
URLs	5

Total Core Research Searches	
Total Searches	240
Time/Result (min)	2
Total Search Time (min)	480
Total Search Time (hours)	8
Hours / day searching	8
Total Search Time (days)	1

## Findings

We found that:

1. The **weighting of the scores** was appropriate without adjustment
2. There were some **irrelevant results** coming from News Sources, Blogs and Social Media, different geographic jurisdictions, as well as from items where **queries were weaker**
3. We were receiving **many duplicate results**

## Improvements to Methodology

As a result of these findings, we:

- **Increased items to four per potentially prohibited item category** to increase item spread, and reduced the number of URLs per query to **12 to compensate**
- Tailored search queries to **emphasise UK relevance**
- Collected a **random sample of webpages from across a deeper set** of Search Engine Results Pages

# In Phase 2, we carried out the fieldwork, collecting and analysing data to inform overall research findings

## 2.0 Execute fieldwork to assess prevalence of sale and supply of potentially prohibited items on Search Services across ~4.6k URLs

2.1

### Data Collection

- Collected and scored ~4.6k webpages over 3 weeks, covering Search Service, Product, potentially prohibited item Category and Type, and Query Type
- Presented interim findings to Ofcom project team each week to pressure test presentation methods and socialise early findings

2.2

### Data Analysis and Quality Assurance

- Cleaned dataset, removing any Personally Identifying Information, identifying and remedying inconsistencies across scorers
- Conducted quantitative and qualitative analysis on dataset, across Controlled and Categorical variables

2.3

### Additional Analyses

- Used ~500 additional webpages to examine DuckDuckGo and additional platform functionalities such as Autocomplete, Related Search and Sponsored Results
- Conducted functional analyses of Bing and Google Search Engine Results Pages

# We scored 4,608 webpages using a robust, human-first approach which has the potential to be automated at scale

## Process Flow

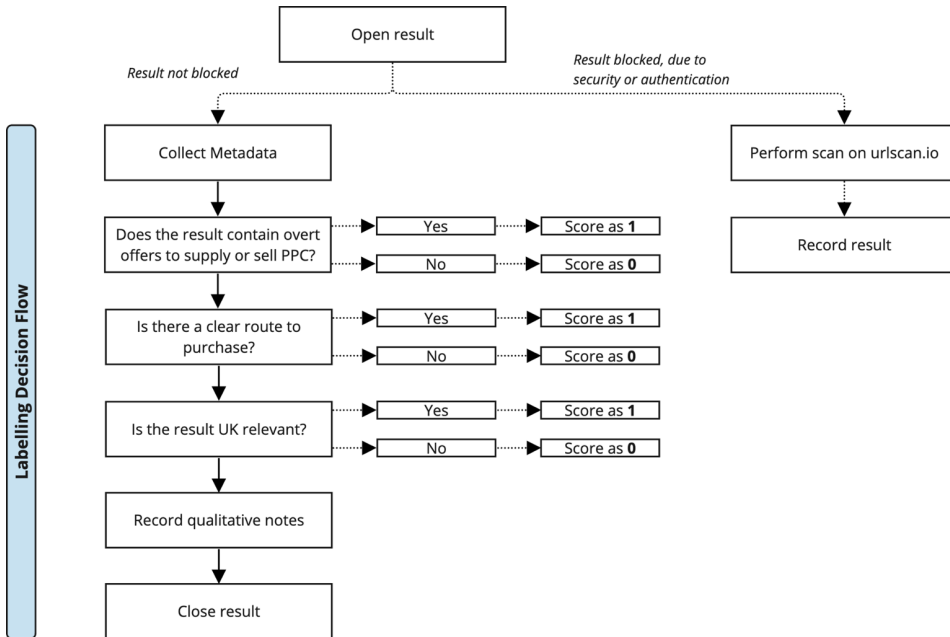


Figure 2: Labelling and scoring results

## Quality Assurance

- The flow we designed to record insights ensured that there was **consistency to how results were being captured**.
- The creation of a quantifiable score also facilitated analysis of prevalence **across difference potentially prohibited items** and categories as well as **search services, products and queries**.
- **Qualitative notes** – outlined in this section – were also taken for further **insights and evidence** gathering.

## Takeaways

- We took a **human-first approach** to this research project to prioritise **richer qualitative insights**, and to observe **platform Terms of Service**
- However, our approach may be **adapted for automated scoring** across much **larger datasets**, based on our rigorous scoring methodology

# Scoring was assigned according to three mutually exclusive factors

Increasing likelihood that potentially prohibited items can be sold to a UK resident



**A score of 3 indicates** that there are reasonable grounds to believe a **potentially prohibited item is offered for purchase**, accompanied by an **apparent route to purchase** for someone accessing the internet **from the UK**.

**A score of 2 indicates** that there are **reasonable grounds** to believe a **potentially prohibited item is offered for purchase**, accompanied by an **apparent route to purchase**.

**A score of 1 indicates** that there are **reasonable grounds** to believe a **potentially prohibited item is offered for purchase**.

**A score of 0 indicates** that there are **no reasonable grounds** to believe a **potentially prohibited item is offered for purchase**.

**Note:** For items where the presence of potentially prohibited items was more challenging to ascertain, researchers took a cautious approach to scoring it is as available which may have resulted in a lower prevalence scores.

# We collected additional data in three key categories to provide greater nuance to results and scoring

These variables were decided in consultation with Ofcom, and have been captured in the Further Insights section of this report (pp.34-40)

## Guidance

Results that contain instructions on how to buy or make potentially prohibited items.

These **could be retailers or e-commerce sites**, and **could be selected for a website that had been classified as meeting criteria to contain potentially prohibited content**.

### For example:

A blog or forum post containing instructions on **where or how to buy cannabis in the UK**, but without an actual product for sale.

## Other potentially prohibited items

Results containing the **presence of other potentially prohibited items as products**.

This was according to legislation available, without additional searching. This **could be selected for a website that had been classified as meeting criteria to contain potentially prohibited content**.

### For example:

An illegitimate retailer of Controlled Drugs, found under **Cocaine**, also advertising **Methamphetamine** or **Benzodiazepines**.

## Obfuscation

Results that we believe referred to a potentially prohibited item, but used a deliberately obscuring or misleading name. This could be selected for a website that had been classified as meeting criteria to contain potentially prohibited content.

### For example:

Referring to **GHB** as **Wheel Cleaner** would not return a positive potentially prohibited content score (unless GHB was mentioned); it would be **obfuscation**.

# These additional areas were investigated with smaller sample sizes and compared with the main research results

	Research Question	Methodology	Indicators
<b>DuckDuckGo</b> *	Are potentially prohibited items <b>more likely to be prevalent on another search service?</b>	<ul style="list-style-type: none"> <li>Assessed <b>~300 webpages</b></li> <li>Used the <b>highest risk potentially prohibited items</b> and search products from main research</li> </ul>	<ul style="list-style-type: none"> <li>Examining number of webpages meeting criteria for containing potentially prohibited content</li> </ul>
<b>Autocomplete and Related Searches</b>	Do <b>query-based search widgets</b> , such as Autocomplete and Related Search, play a role in <b>surfacing potentially prohibited items</b> for users?	<ul style="list-style-type: none"> <li>Assessed <b>~200 'responses'</b> for each feature on Google/Bing</li> <li>Used the <b>highest risk potentially prohibited items</b> and search products from main research</li> </ul>	<ul style="list-style-type: none"> <li>Examining <b>how many autocomplete/related search 'responses'</b> were generated for queries</li> </ul>
<b>Sponsored Results</b>	Do <b>results-based widgets</b> , such as sponsored results, play a role in <b>surfacing potentially prohibited items</b> for users?	<ul style="list-style-type: none"> <li>Assessed <b>~250 webpages</b> across Google/Bing</li> <li>Used the <b>highest risk potentially prohibited items</b> and search products from main research</li> </ul>	<ul style="list-style-type: none"> <li>Examining <b>how many sponsored results were returned</b>, and how many of these returned webpages meeting criteria</li> </ul>

\*Though there is a relationship between DuckDuckGo and Microsoft, there is little information in the public domain to support assumptions into how this may affect the results of this research.

# Quality assurance was embedded throughout the research methodology to increase confidence levels in findings

1

## URL Sampling Strategy

- **Evenly distributing random sampling** across the first three results pages to mitigate against duplicate URLs and strong search engine optimisation (SEO)
- **Using three unique queries** across each potentially prohibited items increase the breadth and representativeness of the analysis.
- **Sampling across four search products** (text, image, shopping, video) to account for various search functionalities.

2

## Research Pilot + Validation Workshop

- **Explaining the proposed data collection strategy** to Ofcom based on pilot research
- **Iterating our chosen variables**, through collaboration with Ofcom
- Changes made following the workshop included: **revising the shortlist** of potentially prohibited items, **revisiting search queries** to include more coded language, **adjusting quantitative metrics** and **adding additional metadata categories** for scoring.

3

## Dataset Validation Exercise

- **Systematically spot-checking** the final dataset for consistent scoring across all scorers
- Making **rules-based adjustments to scoring**, ensuring key categories and variables where mutually exclusive
- Documenting **duplicate URLs**
- Cleaning the data for **personally identifiable information**
- Revisiting **edge cases** and results flagged as ambiguous results

# Mitigations to prevent duplicate webpage analysis from the same query were used, but these were limited to individual queries to reflect user journeys

- 21.7% of all webpages were duplicates
- 28.6% of websites meeting criteria were duplicates
- PUBLIC factored against duplicate webpages across **one user journey** through sampling randomly across a deeper set of search results page tabs
- We have not removed duplicate webpages across **different user journeys**
  - i.e. the fact that a user may encounter the same webpage across different queries should not be counted against in an estimate of prevalence of potentially prohibited items

For all webpages

No. of URL counts    No. of webpages

Unique URL	3683
2 counts	234
3 counts	62
4 counts	15
5 counts	3
6 counts	2
7 counts	0
8 counts	0
9 counts	0
10 counts	0

For webpages meeting criteria

No. of URL counts    No. of webpages

Unique URL	142
2 counts	17
3 counts	5
4 counts	2
5 counts	0
6 counts	0
7 counts	0
8 counts	0
9 counts	0
10 counts	0



## While this work provides evidence of the risk of potentially prohibited content on search services, it has the following limitations

### Sample Size and Interpretation

The small sample size **does not lend itself to making generalized statements** on the basis of the differences, they should rather be used to **direct further investigation**.

To gauge if there are meaningful differences between search services, categories, products or items, **counts should be interpreted in relation to the total sample size** (or as a percentage).

Since the research is limited to four potentially prohibited items per category, findings should not be interpreted as representative, but rather as a **proxy to gauge risk and inform areas for future research**.

### Reproducibility

This research represents a **specific point in time** and should be seen as a **snapshot of search results** during the fieldwork. Given that some webpages that were classified as containing potentially prohibited content were only available for a short time period, it is recommended that future research includes **increased documentation** to mitigate against situations where relevant content cannot be re-accessed post-hoc.

The process of scoring requires **human judgement, which makes the scores subjective measurements**. Post-hoc corrections, dataset validation, and having clear guidance and training available to scorers mitigates this risk.

An **intercoder agreement exercise** is recommended to quantify the extent to which scores are internally consistent across scorers. This measurement is also known as the 'cronbach's alpha'.

### Scalability and Future Research

Finding an appropriate **proxy for geographic relevance of a specific web page is challenging**. Future research should **revise the scoring** to reflect this, for example by dedicating more resources to investigating the georelevance.

To improve the reliability of the analysis an **increased number of webpages is necessary**. These should be allocated to increase the number of **potentially prohibited items per category**, as well as increasing the **number of webpages per query**.

To advance this research and increase its validity, it is critical to **understand and represent the user journey** for those seeking to purchase potentially prohibited items using search services in the UK.

# Summary of Findings

*(Pages 26-36)*

In this section we summarise our high level insights in the following order:

1. By potentially prohibited item **category** and **item** (RQ1, RQ2, RQ5)
2. By **search product** against search service (RQ1, RQ2, RQ4)
3. By **search service** against potentially prohibited items (RQ1, RQ2, RQ4)
4. By **query type** against search service (RQ5)

# PUBLIC has included some guidance on reading and interpreting the findings of this research project

## Key considerations

- Webpages that score a 3 have been “classified as meeting the criteria for containing potentially prohibited content”.
- This is the phrasing that should be used when referring to all results. For example:
  - “The text product had the highest number of webpages classified as meeting the criteria for containing potentially prohibited content”.
- Potentially prohibited items and potentially prohibited content are not the same. Potentially prohibited content refers to content found on a webpage which suggests the sale or supply of potentially prohibited items. Results should not be referred to as having prevalence of potentially prohibited items.
- Whilst we have included depictions of results down to the query level (i.e. 12 webpages) for completeness, these specific results should not be used to infer trends due to small sample sizes.

## Scoring Criteria

Webpages scoring a 3 have met the following criteria:

- There is a **seeming presence of the potentially prohibited item in question** on the site landing page, as a product, without clicking further into the webpage (Score = 1)
- There is a **route to purchase present**, by which the user might progress a transaction to the point of goods or currency exchanging hands (Score = 2)
- The transaction seems possible for **someone residing in the UK** (Score = 3)

# We have found that potentially prohibited items are being offered for sale 'one-click' away from the search results page

## Research Question

**RQ1**

Is content related to the sale or supply of potentially prohibited content accessible 'one click' away from the search service results page?

**RQ2**

What is the prevalence of potentially prohibited content within search results delivered by search services?

**RQ3**

Does functionality on search services play a role in surfacing potentially prohibited content, and if so, how?

**RQ4**

How, if at all, do answers to the above differ between search services, and their products?

**RQ5**

What, if any, patterns are there in the potentially prohibited content that is accessible from search services?

## Findings

Research indicates that potentially prohibited content **is offered for sale 'one click' away from search engine results page**. This varies based on the type of potentially prohibited item, the search service, the search product and the query used to search but all items were accessible.

Across the full dataset, there were 199 webpages classified as meeting the criteria for containing potentially prohibited content.

Autocomplete, search engine results pages, search products, and related and sponsored searches were identified as **functionalities that play roles** in surfacing potentially prohibited content.

- At an aggregated level, there is **no meaningful difference between the number of webpages meeting criteria across Google and Bing, though there are differences across different search products.**
- **Text search has notably more webpages meeting criteria** than image, shopping or video search. On this product, the number of webpages meeting criteria is 36.5% higher on Bing than on Google.

**The category 'Psychoactive substances'** returned the most webpages meeting criteria, whilst **GHB and Cannabis** returned the most webpages meeting criteria among the individual potentially prohibited items.

# Full Results: Psychoactive Substances

Count of webpages that meet the criteria for containing potentially prohibited content

Item	Query	Bing					Google					Grand Total
		Text	Image	Shopping	Video	Total	Text	Image	Shopping	Video	Total	
GHB	Primary	5	0	0	0	5	3	0	0	3	6	11
	Secondary	1	1	0	0	2	1	1	0	7	9	11
	Extended	0	2	0	0	2	3	1	0	1	5	7
	<b>Total</b>	<b>6</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>9</b>	<b>7</b>	<b>2</b>	<b>0</b>	<b>11</b>	<b>20</b>	<b>29</b>
Mephedrone	Primary	0	0	0	0	0	1	0	0	1	2	2
	Secondary	2	3	0	0	5	1	0	0	0	1	6
	Extended	4	1	0	0	5	0	0	0	0	0	5
	<b>Total</b>	<b>6</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>10</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>3</b>	<b>13</b>
Spice	Primary	1	2	0	0	3		0	0	3	3	6
	Secondary	0	0	0	0	0	0	1	0	0	1	1
	Extended	2	1	0	0	3	3	0	0	0	3	6
	<b>Total</b>	<b>3</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>6</b>	<b>3</b>	<b>1</b>	<b>0</b>	<b>3</b>	<b>7</b>	<b>13</b>
MXE	Primary	0	3	0	3	6	0	0	0	0	0	6
	Secondary	0	0	0	0	0	0	0	0	0	0	0
	Extended	0	0	0	0	0	2	0	0	0	2	2
	<b>Total</b>	<b>0</b>	<b>3</b>	<b>0</b>	<b>3</b>	<b>6</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>8</b>
<b>Psychoactive Substances Total</b>		<b>15</b>	<b>13</b>	<b>0</b>	<b>3</b>	<b>31</b>	<b>14</b>	<b>3</b>	<b>0</b>	<b>15</b>	<b>32</b>	<b>63</b>

# Full Results: Controlled Drugs

Count of webpages that meet the criteria for containing potentially prohibited content

Item	Query	Bing					Google					Grand Total
		Text	Image	Shopping	Video	Total	Text	Image	Shopping	Video	Total	
Cannabis	Primary	5	1	0	0	6	0	0	1	0	1	7
	Secondary	3	0	0	0	3	1	0	0	0	1	4
	Extended	3	0	0	0	3	7	1	0	1	9	12
	<b>Total</b>	<b>11</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>12</b>	<b>8</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>11</b>	<b>23</b>
Codeine	Primary	0	0	0	0	0	0	0	0	5	5	5
	Secondary	2	0	0	0	2	0	0	0	0	0	2
	Extended	0	0	0	0	0	7	0	0	0	7	7
	<b>Total</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>7</b>	<b>0</b>	<b>0</b>	<b>5</b>	<b>12</b>	<b>14</b>
Ecstasy	Primary	3	0	0	0	3	0	0	0	0	0	3
	Secondary	3	0	0	0	3	0	0	0	0	0	3
	Extended	4	0	0	0	4	0	0	0	0	0	4
	<b>Total</b>	<b>10</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>10</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>10</b>
Cocaine	Primary	1	0	0	0	1	0	0	0	0	0	1
	Secondary	3	0	0	0	3	1	1	0	0	2	5
	Extended	0	0	0	0	0	3	0	0	0	3	3
	<b>Total</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>	<b>4</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>5</b>	<b>9</b>
<b>Controlled Drugs Total</b>		<b>27</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>28</b>	<b>19</b>	<b>2</b>	<b>1</b>	<b>6</b>	<b>28</b>	<b>56</b>

# Full Results: Knives and Bladed Weapons

Count of webpages that meet the criteria for containing potentially prohibited content

Item	Query	Bing					Google					Grand Total
		Text	Image	Shopping	Video	Total	Text	Image	Shopping	Video	Total	
Baton*	Primary	4	1	0	0	5	0	0	3	0	3	8
	Secondary	1	0	0	0	1	2	0	1	0	3	4
	Extended	0	0	1	0	1	2	2	2	0	6	7
	<b>Total</b>	<b>5</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>7</b>	<b>4</b>	<b>2</b>	<b>6</b>	<b>0</b>	<b>12</b>	<b>19</b>
Disguised Knife	Primary	1	1	2	0	4	1	1	1	0	3	7
	Secondary	2	0	2	0	4	2	0	0	0	2	6
	Extended	0	0	1	0	1	1	2	1	0	4	5
	<b>Total</b>	<b>3</b>	<b>1</b>	<b>5</b>	<b>0</b>	<b>9</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>0</b>	<b>9</b>	<b>18</b>
Butterfly Knife	Primary	1	0	0	0	1	0	3	0	0	3	4
	Secondary	2	2	0	0	4	2	2	0	0	4	8
	Extended	3	1	0	0	4	0	1	0	0	1	5
	<b>Total</b>	<b>6</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>9</b>	<b>2</b>	<b>6</b>	<b>0</b>	<b>0</b>	<b>8</b>	<b>17</b>
Zombie Knife	Primary	0	3	0	3	6	0	0	0	0	0	6
	Secondary	0	0	0	0	0	0	0	0	0	0	0
	Extended	0	0	0	0	0	0	0	1	0	1	1
	<b>Total</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>1</b>
<b>Knives and Bladed Weapons Total</b>		<b>14</b>	<b>5</b>	<b>6</b>	<b>0</b>	<b>25</b>	<b>10</b>	<b>11</b>	<b>9</b>	<b>0</b>	<b>30</b>	<b>55</b>

\*Baton has been included in this category as it is covered by similar guidance and legislation to other knives and bladed weapons





# Full Results: Firearms

Count of webpages that meet the criteria for containing potentially prohibited content

Item	Query	Bing					Google					Grand Total
		Text	Image	Shopping	Video	Total	Text	Image	Shopping	Video	Total	
Stun guns	Primary	0	0	0	0	0	1	0	0	0	1	1
	Secondary	0	2	0	0	2	1	0	0	1	2	4
	Extended	4	0	0	0	4	2	0	0	0	2	6
	<b>Total</b>	<b>4</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>6</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>5</b>	<b>11</b>
Handgun	Primary	0	0	0	0	0	0	1	0	0	1	1
	Secondary	1	0	0	0	1	0	3	0	0	3	4
	Extended	0	0	0	0	0	0	1	0	0	1	1
	<b>Total</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>5</b>	<b>0</b>	<b>0</b>	<b>5</b>	<b>6</b>
Realistic imitation firearm	Primary	0	0	0	0	0	0	0	0	0	0	0
	Secondary	2	0	0	0	2	0	0	0	0	0	2
	Extended	1	0	0	0	1	0	1	0	0	1	2
	<b>Total</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>4</b>
Ammunition with noxious substances	Primary	0	0	0	0	0	0	0	0	0	0	0
	Secondary	3	0	0	0	3	0	0	0	0	0	3
	Extended	1	0	0	0	1	0	0	0	0	0	1
	<b>Total</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>
<b>Firearms Total</b>		<b>12</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>14</b>	<b>4</b>	<b>6</b>	<b>0</b>	<b>1</b>	<b>11</b>	<b>25</b>



# Psychoactive substances returned the most webpages meeting the criteria, followed by Controlled Drugs and Bladed Weapons





Prohibited Category	Potentially Prohibited item	Prevalence (number of websites meeting criteria)	
 <b>Psychoactive Substances</b>	GHB	29/288	63/1152
	Mephedrone	13/288	
	Spice	13/288	
	MXE	8/288	
 <b>Controlled Drugs</b>	Cannabis	23/288	56/1152
	Codeine	14/288	
	Ecstasy	10/288	
	Cocaine	9/288	
 <b>Knives and Bladed Weapons</b>	Baton	19/288	55/1152
	Disguised knife	18/288	
	Butterfly knife	17/288	
	Zombie knife	1/288	
 <b>Firearms</b>	Stun guns	11/288	25/1152
	Handguns	6/288	
	Realistic imitation firearm	4/288	
	Ammunition with noxious substances	4/288	

## Findings

- Every potentially prohibited item researched was available on at least one occasion, which suggests that, in general, **potentially prohibited content is present 'one-click' from the search engine results page.**
- **Psychoactive substances had the most websites classified as meeting the criteria for containing potentially prohibited content** across all categories researched, while **firearms had the least.**
- **PUBLIC used items with the highest number of webpages meeting criteria to investigate additional hypotheses** across DuckDuckGo and other search service products.

**NOTE:** These results are aggregated across searches all search products and services. While this may indicate risk of particular potentially prohibited items, further research is required to establish firm trends.

# Text search returned the highest number of webpages meeting criteria versus other search products, such as shopping and video

Search Product	Search Service	Prevalence (number of incidences found)	
 <b>Text Search</b>	Google	47/576	115/1152
	Bing	68/576	
 <b>Image Search</b>	Google	22/576	43/1152
	Bing	21/576	
 <b>Video Search</b>	Google	22/576	25/1152
	Bing	3/576	
 <b>Shopping Search</b>	Google	10/576	16/1152
	Bing	6/576	

## Findings

- **Text search** had the highest number of webpages meeting the criteria for containing potentially prohibited content.
- **Bing's text search returned 36.5% more webpages meeting criteria than Google's.**
- **Google's video search returned 19 more webpages meeting criteria than Bing's.** There is a **difference in the functionality of Google and Bing's video search** that is [explored in the Functional Analysis section of this report](#).
- Comparatively **fewer returned webpages meeting criteria when using video and shopping search** suggest **lower risk** on these products.



**NOTE:** These results are aggregated across searches for all potentially prohibited items. While this may indicate risk of search products, further research is required to establish firm trends.

# Google and Bing returned similar numbers of webpages meeting criteria for each of the potentially prohibited item categories

Search Service

Prohibited Categories and Item with prevalence

Google

Psychoactive Substances		Controlled Drugs		Knives and Bladed Weapons		Firearms		101/2304
GHB	20	Cannabis	11	Baton	12	Stun guns	5	
Mephedrone	3	Codeine	12	Disguised knife	9	Handguns	5	
Spice	7	Ecstasy	0	Butterfly knife	8	Realistic imitation firearm	1	
MXE	2	Cocaine	5	Zombie knife	1	Ammunition with noxious substances	0	
32/576		28/576		30/576		11/576		

Bing

Psychoactive Substances		Controlled Drugs		Knives and Bladed Weapons		Firearms		98/2304
GHB	9	Cannabis	12	Baton	7	Stun guns	6	
Mephedrone	10	Codeine	2	Disguised knife	9	Handguns	1	
Spice	6	Ecstasy	10	Butterfly knife	9	Realistic imitation firearm	3	
MXE	6	Cocaine	4	Zombie knife	0	Ammunition with noxious substances	4	
31/576		28/576		25/576		14/576		

Findings

- The number of websites meeting criteria returned across different potentially prohibited item categories is **comparable across Google and Bing.**
- **However some variations can be found across individual potentially prohibited items.**
  - The **sample size** of URLs analysed at this level is **sufficiently small** that **implications can not be drawn.**

**NOTE:** These results are aggregated across search products. While this may indicate risk of search service, further research is required to establish firm trends.

# Using an extended query returned the greatest number of websites meeting criteria, a finding that could inform further interrogation

Query Type	Search Service	Prevalence (number of websites meeting criteria)	
Primary Query	Google	28/768	62/1536
	Bing	34/768	
Secondary Query	Google	28/768	63/1536
	Bing	35/768	
Extended Query	Google	45/768	74/1536
	Bing	29/768	

## Findings

- The number of websites meeting criteria was **higher when the extended query was used.**
  - This is likely to be as a result of the **more targeted slang or coded language** used in this query and the inclusion of **additional terms related to purchasing of items.**
  - This finding **may have implications for how alternative services** that could be used to search - especially for **AI chatbots** - may **surface potentially prohibited content.**
- This finding may be primarily driven by Google's results. The number of websites meeting criteria is actually **43% lower on Bing when using the extended query.**
  - This finding may require **further analysis** to validate.
- Researcher discretion determined the exact language of queries. This **approach may be refined** in further research.

# Further insights

*Cross-variable analysis*

*(Pages 37-48)*

In this section we conduct a broader set of analyses of the dataset. These include examining:

1. Websites that meet criteria for containing potentially prohibited content, against their position in the **Search Service Page Index**
2. **Additional categorical variables** that were collected, such as Obfuscation, Other types of potentially prohibited items and Guidance
3. Different **Routes to Purchase** found in websites meeting criteria
4. An additional Search Service, **DuckDuckGo**
5. **Qualitative characteristics** of websites meeting criteria

# Webpages meeting criteria tend to appear higher up on the search engine result page

## Text search

Placement in search engine results page	Bing	Google	Grand Total
1-10	27	28	55
11-20	21	7	28
21-30	20	12	32
<b>Grand Total</b>	<b>68</b>	<b>47</b>	<b>115</b>

## Image search

Placement in search engine results page	Bing	Google	Grand Total
1-10	7	10	17
11-20	10	5	15
21-30	4	7	11
<b>Grand Total</b>	<b>21</b>	<b>22</b>	<b>43</b>

*Video and shopping search have a lower total number of incidences and have been excluded from this analysis due to small sample size*

## Total across all search products

Placement in search engine results page	Bing	Google	Grand Total
1-10	39	55	94
11-20	34	24	58
21-30	25	22	47
<b>Grand Total</b>	<b>68</b>	<b>47</b>	<b>115</b>

## Takeaways

- Many of these sites might make use of **SEO** to ensure their sites appear high up on the Search Index.
- Only the top 30 results were analysed. **Users may go to results deeper on the results lists to attempt a purchase. This research did not** investigate user behaviours when searching for potentially prohibited items.

# We collected additional categorical data, which can enrich our understanding of the design of websites that meet criteria

Route to possible potentially prohibited item	Bing	Google	Grand Total
Obfuscation	13	26	39
Guidance	16	28	44
Other potentially prohibited items	249	120	369
<b>Grand Total</b>	<b>68</b>	<b>47</b>	<b>115</b>

**Obfuscation:** Where it appears that a potentially prohibited item is being sold, but it is not actually listed under a the name or label of a potentially prohibited item.

**Other potentially prohibited items:** Where types of items identified in the longlist of all potentially prohibited items but not the item searched for were also present on the page. For example, searching for cocaine and finding other Controlled Drugs also on the same page.

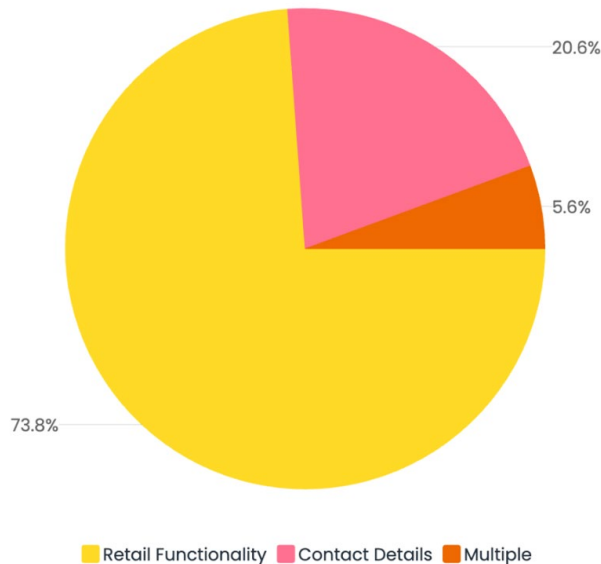
**Guidance:** Speaks to blogs or forums that instruct or guide users on how to purchase the potentially prohibited item.

## Takeaways

- **Obfuscation** was the least common variable observed. This may be because obfuscation of the item being sold reduces user visibility on a Search Results Page. Instead, bad actors appear to prioritise SEO and product keywords to promote their sites in the search index.
- **Other types of potentially prohibited items** were the most common variable found; this was often identified for websites meeting criteria in the Controlled Drugs category. This is characteristic of illegitimate online pharmacies or 'research chemical' sites that advertise many different Controlled Drugs items at one time.
- **Guidance** was also not recorded frequently. This might be explained by the transactional nature of queries used. Where it was found, it was usually on blogs and forums. We did not come across any guidance regarding accessing the Dark Web in the research.

# On websites containing potentially prohibited content, sale was mostly enabled by typical e-commerce functionality

## Route to purchase for available potentially prohibited items



- **Route to Purchase:** Whether the result contains an apparent route to purchase
- **Retail Functionality:** Features such as 'Add to Cart', Shopping Baskets, or quantity selectors.
- **Contact Details:** Details such as email addresses, phone numbers, and online identifiers (e.g., social media handles).

## Takeaways

- **Conventional retail functionality** was the most common route to purchase by a significant margin (53%).
- **Where retail functionality was leveraged** as the means to acquire a potentially prohibited item, the **currency required to purchase was GBP on 73.8%** of occasions.
- Displaying of **contact details** to execute a purchase or **multiple forms of routes to purchase** were also identified.

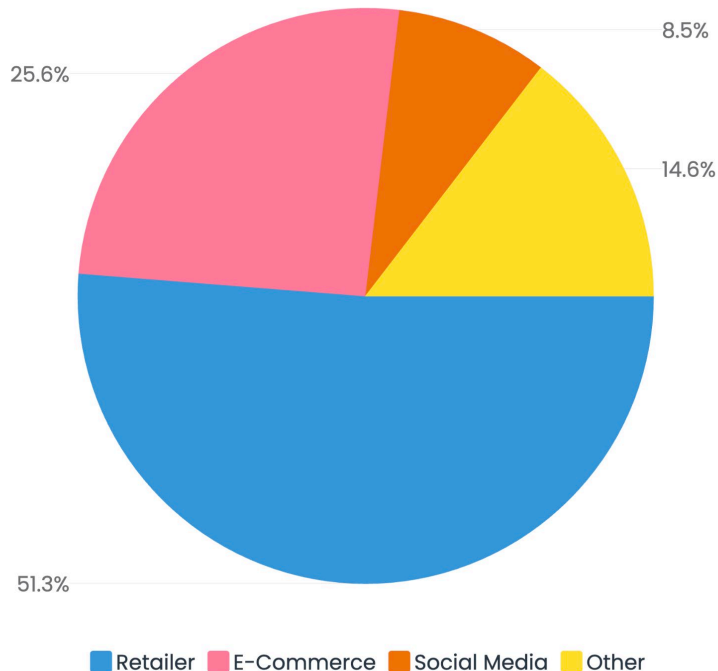
*Note: If sites contained warnings such as 'Do not order without a valid prescription'; or required a user to login in order to access Retail Functionality features, we did not count this as a Route to Purchase.*

*Third party URLs were not considered as Contact Details. Some URLs contributing to this figure may be duplicates (from the same site) encountered across multiple queries.*



# Webpages meeting criteria were predominantly retail and e-commerce sites, followed by a smaller number of social media sites

Type of website meeting criteria







## Findings

- A **majority of the websites meeting criteria for containing potentially prohibited content were Retailer (102 websites) or e-commerce (51 websites).**
- There were **17 social media webpages that met criteria.**
- Other webpages that met criteria for containing potentially prohibited content included a variety of **nefarious uses of other types of websites.**

# We found a comparable number of webpages meeting criteria on DuckDuckGo, a search service which prioritises user privacy

After preliminary research suggested potential higher prevalence on DuckDuckGo, an **alternative search service focused on user privacy**, we investigated the service using the same methodology as other services, but a smaller number of URLs.

Prohibited Category	Potentially prohibited item – product	Prevalence (number of websites meeting criteria)	
 <b>Psychoactive Substances</b>	GHB - Text	8/36	15/72
	GHB - Image	7/36	
 <b>Controlled Drugs</b>	Cannabis - Text	23/36	36/72
	Cannabis - Image	13/36	
 <b>Knives and Bladed Weapons</b>	Baton - Text	2/36	2/72
	Baton - Image	0/36	
 <b>Firearms</b>	Stun guns - Text	2/36	5/72
	Stun guns - Image	3/36	

## Findings

- We examined **288 URLs from DuckDuckGo** taking the highest scoring items and search service products from the original research.
- The number of webpages found was **comparable to Google and Bing for these search items and products specifically**.
- On DuckDuckGo, around **20% of results** met the criteria for containing potentially prohibited content
- Across Google and Bing, **for the same metrics**, around **21%** contained a route to purchase potentially prohibited items.
- The **sample size** for this is **not large enough to draw firm insights** but indicative that results on DuckDuckGo are broadly comparable to Google and Bing.

# We could distinguish between a webpage containing potentially prohibited content related to Controlled Drugs and Psychoactive Substances versus one containing potentially prohibited content related to Bladed Weapons and Firearms

## Bladed Weapons and Firearms

- There were **many potentially prohibited items present, and for sale**, often with a **valid route to purchase**.
- However, Bladed Weapons and Firearms are **not consistently regulated across the globe**. This meant the point of difference between a result scoring a '2', and a result scoring a '3', was ascertaining if the seller was **likely to ship to the United Kingdom**.
- We used several indicating factors to make this assessment during this research. However, we recommend further investigation into this for future assessments.
- Webpages that met criteria for containing potentially prohibited content were broadly **international e-commerce** or **retail sites**.

## Controlled Drugs and Psychoactive Substances

- In this grouping, it was easier to distinguish which results were the highest risk.
- This meant that when a site appeared to be selling or supplying the items, it was unlikely that there was a jurisdiction in which this supply would be legal.
- **Codeine** introduced some difficulty in classification as there are some legal online routes to purchase (i.e. with a valid, checked prescription). Similarly, one of the queries for GHB mentions GBL which can be sold legally as wheel cleaner.
- We found two main types of site which were often classified as meeting the full scoring criteria: **retail sites, and** sites filled with seemingly web-scraped text containing many SEO keywords.

# We used several factors to assess likelihood of illegal shipping to the UK

*However, further research in this area may benefit from collaboration with law enforcement to set rigorous assessment criteria for this particular risk factor*

**Post to me**  
 Free P&P on each additional eligible item you buy from id-paintball.  
 Item location: **Berlin, Germany**  
 Posts to: Worldwide  
 Excludes: Asia, Bulgaria, Cyprus, Finland, France, Greece, Lithuania, Middle East, Poland, Romania, Russian Federation, Slovakia, South East Asia, Spain, Sweden, Ukraine  
 Quantity:  Change country:  [Get rates](#)

★★★★★ J\*\*\*e

Wow, this is a nice knife, great price, and fast delivery, honest seller, highly recommend seller, nice products, A+++



UK

Postage: EUR 20.17 (approx £17.59) Standard Delivery. [See details](#)  
 Located in: Berlin, Germany

Delivery: Estimated between Wed, 17 May and Thu, 25 May   
 Seller dispatches within 1 day after receiving cleared payment.

Returns: 1 month return. Buyer pays for return postage. [See details](#)

Postage: **Doesn't post to United Kingdom.** [See details](#)  
 Located in: San Juan, Puerto Rico, United States

!!! We do not ship this item to United Kingdom !!!

Continental US shipping only. Individual state restrictions apply. We are unable to ship internationally, including Mexico and Canada.

## Factors which increased likelihood included:

- UK Currency
- UK site domain
- Price or timings for shipping to the UK
- Reviews from buyers originating in the UK
- Disclaimers

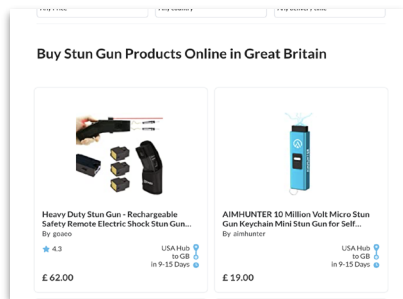
## Factors which decreased likelihood included:

- Non-UK Currency
- Warnings against ordering to the UK
- Inclusion of UK in 'Excluded countries' lists
- Specifying geographic delivery (i.e. US Continental)

# Some e-commerce sites appear to be selling multiple types of potentially prohibited items across different global jurisdictions

## Type of Site

E-commerce



**NOTE:** Some images have been redacted from the public-facing version of this report.

## Potentially prohibited item categories

Knives and Bladed Weapons  
Firearms

## Observations

These sites often sell **many variations** of the same products, from several manufacturers, though items are **not branded**. Items may be **sold wholesale**. The sites or their sellers use many **targeted SEO keywords** to describe items. They appear to **ship globally**, might label products with multiple currencies, and often have **estimated global shipping times and costs**.

## Takeaways

- Some sites meeting the full scoring criteria contained multiple potentially prohibited items, and appeared multiple times across queries.
- Some of these sites might be unreliable or scam websites; further research is needed to understand the overlap between prevalence of potentially prohibited content and financial crime

# Some retailers appear to only sell potentially prohibited items, particularly Controlled drugs and Psychoactive substances

PUBLIC

Type of Site

Retailer

Potentially prohibited item categories

Controlled Drugs

Psychoactive Substances

## Observations

These sites have a **similar user experience and interface** to legitimate retailers, but appear to sell **only controlled substances**. They feature chatbots, shopping baskets, quantity widgets and **discounts**. They often feature many **other types of potentially prohibited item** alongside the one searched for. They also feature high use of **SEO keywords and tagging**.

## Takeaways

- Many of these sites may be **scam sites**, but this is difficult to determine without **law enforcement expertise**. Work should be carried out with law enforcement to determine **markers of reliable or verifiable attempts** to sell or supply.
- The domain names of these sites often contained **wording indicating their purpose**, such as ganjashop.co.uk, cannabisbudshop.co.uk, buycocaineonlineinuk.co.uk.

Buy Cocaine in UK  
\$400.00 - \$23,000.00

Unit Measure & Prices: 7 grams, 14 grams, 28 grams, 56 grams, 112 grams, 224 grams, 448 grams, 1000 grams (1kg)

Unit Measure & Prices: Choose an option

Chat

HOME - RESEARCH CHEMICALS  
Buy MDPV Powder online  
\$280.00 - \$1,030.00

Buy MDPV online is one of the main chemicals found in powders sold as 'bath salts'. It is a white crystalline powder in its pure form, but manufacturing impurities often render it from off-white to pale brown.

GRAMS: Choose an option

ADD TO CART

SKU: N/A

Category: Research Chemicals

Tags: buy mdpv online, Buy MDPV Powder online, buy mdpv uk, buy mdpv with credit card, mdpv crystal for sale, mdpv crystalline for sale, mdpv for sale with paypal, mdpv powder for sale, mdpv salt for sale, order mdpv in UK, pure mdpv crystals, where can i buy mdpv crystals, where to buy mdpv powder.

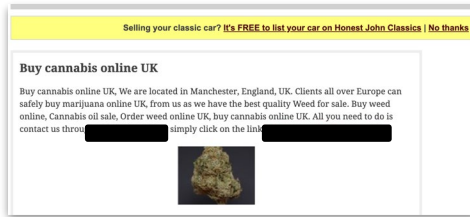
Someone in URRAEEN, Queensland, Australia purchased a  
Magic mushroom grow kit Ecuador XP

-59%

# Some webpages classified as meeting criteria contained web-scraped text relevant to the search query used

## Type of Site

## Sites with web-scraped text



suppressant Dosage Form: Syrup Validity: 2yrs Strength 200mg Drug: 1622-62-44 16 oz, 32 oz List Of Available Cough Syrup: Hi Tech Actavis Promethazine Qualitest Syrup Tusonoxen Syrup Ketof Cough syrup 100ml Rexcof Cough Syrup Green Caraseo syrup Wockhardt Promethazine Cough Syrup Home / Pain Meds \$ 350.00 - \$ 1,500.00 Buy Wockhardt promethazine with codine cough syrup We sell top quality Actavis Promethazine with codine purple cough syrup in large quantities at very affordable prices. Place your order and contact now. Validity: 2 years Strength 200 mg 1 Pint (473mL) = 16 Oz Quantity Cough due to a common cold often does not need to be treated with medicine Ask your doctor or pharmacist about other ways to relieve cough and cold symptoms, such as drinking enough fluids using a humidifier, or saline nose drops/spray. Promethazine cough syrup for sale in Canada. Buy Promethazine DM Syrup online. Taking codine with other medicines and herbal supplements Common questions Related conditions. Diarrhoea and vomiting 10 ways to reduce pain Useful resources. HealthUnlocked: codine forum healthunlocked.com British Pain Society - charity - people living with pain. Actavis Promethazine Codine Cough Syrup SKU: wda. Category: Spice Herbal Essence Tags: Actavis Promethazine Codine Cough Syrup, Actavis Promethazine Codine Cough Syrup Online, Buy Actavis Promethazine Codine Cough Syrup Online \$ 180.00 Tweet Packs Clear Add to cart Add to wishlist Compare Additional Information Reviews (0) Buy Promethazine Syrup Promethazine/codine is a low cost medication used as an antihistamine, a decongestant, and a cough suppressant. Coughing and congestion are lessened or stopped thanks to it. Compared to similar medications, it is more widely used. Buy Actavis Promethazine codine online Buy Actavis Cough syrup a pain reliever and a cough suppressant. Actavis Promethazine codine is an antihistamine. It blocks the effects of the naturally occurring chemical histamine in your body. The combination of codine and prometha treats cold or allergy symptoms such as a runny nose, sneezing, and cough. Buy Wockhardt Online. Secondly, Promethazine codine. Codine is a medicine that is used to relieve pain and to reduce coughing, and codine with a prescription, and without prescription. Moreso, Codine is one type of opioid. It is found in many cough medicine. You can buy Cough medicines with codine at drugstores without a prescription in a Main Menu. Search. Buy Akorn Lean Online 1 1 product; Buy Actavis Promethazine With Codine 1 1 product; Buy Hi Tech Cough Syru Online 1 1 product; Buy Quango Lean Online 1 1 product; Buy Qualitest Lean Online 3 3 products; Buy Tris Cough Syrup Online 1 1 product; Buy Wockhardt Online Buy Zovirax (Aciclovir) Cream online Free Shipping on orders over \$35 Sort by Zovirax Cream (1 x 2g) Out of stock Zovirax Cold Sore Cream has been clinically proven to be able to prevent cold sores from appearing. The cream contains the unique antiviral agent Aciclovir, which prevents the cold sore virus from replicating. The product being recalled actavis codine promethazine cough syrup for sale is Btarna plain (flat-leaf) parsley with a harvest date of Aug. Food and Drug Administration officials and actavis codine

## Potentially prohibited item categories

## Controlled Drugs

## Psychoactive Substances

## Observations

These sites appear to feature **text scraped from across the clear web** targeted around **SEO keywords** relevant to the user search. Often within these there might be **contact details**, such as telephone numbers and email addresses or social media handles for a user, and advertised quantities or types of controlled substances.

## Takeaways

- Often this content would appear as a **post or review on unrelated sites or forums**, or as text entered into an internal search bar.
- Often the **same contact details (email addresses and social media handles)** would appear over multiple different sites, meaning that potentially a **single bad actor** was using malicious SEO tactics to surface their contact details across many pages.

# Functional Analysis

*(Pages 48-60)*

This section explores ways in which the **functionality** of Search Services might impact the user experience and ability to surface webpages classified as meeting criteria, touching on:

- The **user journey** for a Search Service
- The different Search Service **products**
- The functionality and impact of common search service functionality such as **Autocomplete, Related searches and Sponsored results**



# We complemented the quantitative analysis with an understanding of the functionality of search services

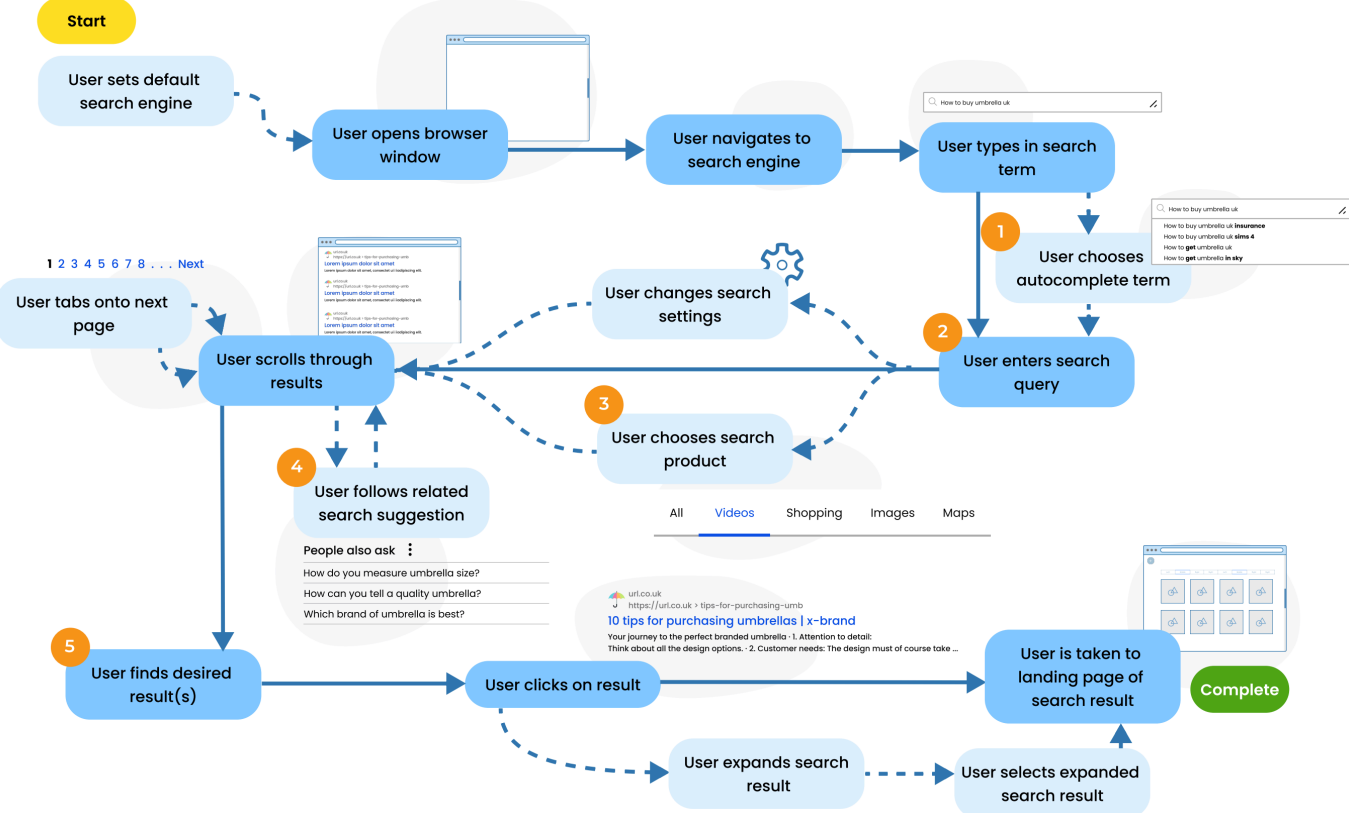
## Objectives

- 1 Explore potential patterns between websites meeting criteria, using quantitative findings and service design
- 2 Explore priority additional **research areas identified by Ofcom**: Autocomplete, Related searches and Sponsored results

## Key Takeaways

1. Search service products return **varying numbers of websites that meet the criteria for containing potentially prohibited content**
2. These results can in part be explained by differences in **functional implementation**
3. Understanding **Autocomplete, Related Searches** and **Sponsored Result** functionalities helps to understand how Search Service content moderation works

# We mapped functionalities in a high-level representative search service user journey



- 1 Autocomplete
- 2 Results Page Analysis
- 3 Search Product
- 4 Related Search
- 5 Sponsored Results

Note. This map was designed using source information from [Microsoft](#) and [Google](#), as well as observations from conducting the fieldwork.

# “Autocomplete” may guide users towards websites that meet criteria but are not covered by content moderation algorithms

## Feature

### Autocomplete/Autosuggest

## Product Description



**Limitation:** We researched Autocomplete across 192 additional queries, assessing only whether results were returned, and not assessing the type of content returned. See full methodology in slides 12 – 25.

**Autocomplete/Autosuggest** is designed to help people complete a search they were intending to do.

According to Google and Microsoft documentation, it is based on **common and trending searches** relevant to the **characters entered by the user**, and is related to location and previous searches (1).

It can be **plain text**, or more **detailed and structured information**, such as **extra illustrative detail about a response**, such as **geographic location or weather** (2).

## Findings

### Summary Statistics

Response	#	%
Suggestions	126/192	66%
N/A	66/192	34%

### Bing/Google

Service	#	%
Google	47/126	37%
Bing	79/126	63%

### Query Type

Query Type	#	%
Root	41/126	33%
Coded	37/126	29%
Extended	48/126	38%

## Trend Elaboration

- **66% of queries** received autocomplete suggestions, and **34% did not**. This might indicate that the autocomplete algorithm proactively removes potentially prohibited items from its suggestions. However **not every query** was covered by these systems; this might indicate that potentially prohibited items are not comprehensively covered by content moderation systems.
- Queries referring to items that are not **completely or universally prohibited**, such as **GBL, stun guns, or batons**, were more likely to receive suggestions.
- Bing delivered **26% more autocomplete suggestions** on queries than Google for **the same set of queries**. The greatest differences arose across **Cannabis and Stun Guns**.

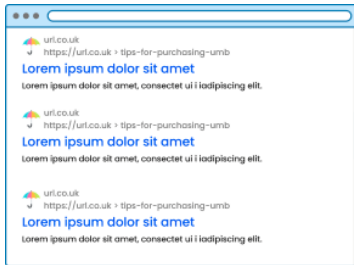
Sources: (1) Google, 2023. [How Google autocomplete works in Search](#), (2) Bing, 2023. [A deeper look at Autosuggest](#)

# The Search Engine Results Page uses combinations of ranking algorithms to surface results to users

## Feature

## Search Results Page

## Product Description



The [Search Engine Results Page](#) is the page that users see after they click to enter their search query (1, 2). The results page differs depending on what search service product the user is engaging with. The most important factors that differed according to search service product, beyond scoring criteria were:

- **How much information** was available at a single glance
- **The type of result** returned (Retail/News etc)
- **How many results** were returned from the same domain

## Findings

## Websites meeting criteria:

Bing/Google: N/A

Source Type: N/A

N/A

Microsoft and Google **did not differ significantly** in the design of their results pages, apart from in the Video page (see slide 56). Both services use automated ranking systems to present the most relevant and useful results. Factors may include: **Meaning, Relevance, Quality, Usability**, and **Context**. [Google](#) provides more information on its systems than Microsoft (3):

- **BERT, a family of language models:** understands how combinations of words express different meanings and intents based on converting words to integers and creating neural networks of words and phrases (4).
- **Exact match domain system:** considers words in domain names
- **Freshness systems:** surfaces 'recent' results
- **Helpful content system:** prioritises original content over traffic-focused content
- **Link analysis:** understands how multiple pages link to one another
- **Neural matching:** links different representations of content across different domains
- **RankBrain:** AI linking words and concepts
- **Reliable information systems:** Surface and demote authoritative and low quality pages
- **Site diversity system:** prevents showing more than two pages from the same site in top results

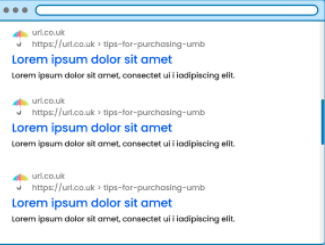
## Trend Elaboration

# Both Google and Bing curate content for users as well as providing users with tools to curate the results page

## Feature

### Curation of the results page

## Product Description



**Search Services** remove results from their page indexes under certain conditions, such as: **legal requests** from government or individuals; **spam**; **sensitive personal information**, including intimate image-based abuse; **adult content**; **illegal pharmaceuticals**; **public service announcements** (1,2).

**Google** also uses **removal-based demotion systems** which demote sites which have received a high volume of Legal or Personal Information removal requests (2).

## Findings

Website meeting criteria: N/A

Bing/Google: N/A

Source Type: N/A

## Trend Elaboration

Different services provide tools for users to customise their search experience at the point of search (1, 2, 3, 4):

- **SafeSearch** is provided by Google and Microsoft, which filter certain results out automatically. Where **Bing** provides levels, **Google** provides an on-off toggle, defaulted to off.
- Google also offers **verbatim searching**, which forces the engine to search for exactly and only the search term.
- Both Google and Bing offer **Date Filtering** of results
- Both **Google** and **Bing** offer advanced search, which allows the user to specify targeted search conditions.

Both services also offer account-level settings for their search browsers, such as Bing's **Kids Mode**, and Google's **Family Link**, which allow parents or guardians to set fixed settings for their children when browsing (5, 6).

Sources: Image: Google, 2023. [Visual elements gallery of Google Search](#), (1) Google, 2023. [A guide to Google Search Ranking Systems](#) (2) Bing, 2023. [How Bing delivers search results.](#) (3) Google, 2023. [Advanced Search](#), (4) Bing, 2023. [Advanced search options](#), (5) Bing, 2023. [Kids Mode](#) (6) Google, 2023. [Family Link](#), \_

# The high number of webpages meeting criteria found through Text search may be attributable to search engine optimisation tactics

PUBLIC

3

## Feature


## Text Search

## PPC Prevalence

115/1152 URLs  
10.0%

This slide builds on the **core research** findings. Methodology on slides 12-25

## Product Description

 url.co.uk  
<https://url.co.uk/tips-for-purchasing-umb>  
**10 tips for purchasing umbrellas | x-brand**  
 Your journey to the perfect branded umbrella - 1. Attention to detail:  
 Think about all the design options. - 2. Customer needs: The design must of course take ...

The Search Engine Results Page differs depending on what search service product the user is engaging with. **Text results** are those based on the textual content of the page. The visual elements usually include an **attribution, title link and snippet**. It may also include additional visual elements such as **rich attributes** (product information such as price or availability), or sitelink groups; this may vary depending on device used and what was searched for. Results found during this project typically did not have many rich attributes or sitelink groups (1, 2).

## Findings

### Websites meeting criteria

Cut of data	#	%
Text/ <b>Total PPC</b>	115/195	60%
Text/ <b>Text URLs</b>	115/1152	10%

### Bing/Google

Service	#	%
<b>Google</b>	47/115	40%
<b>Bing</b>	68/115	59%

### Source Type

Source Type	#	%
<b>Retailer</b>	63/115	55%
<b>E-Commerce</b>	25/115	22%

## Trend Elaboration

- The high number of webpages meeting criteria for containing potentially prohibited content may be explained, in some degree, by **search engine optimisation strategies** used by websites; the language of the research search queries was often reflected **in many different iterations** on the result landing pages, as tags, headlines and captions.
- **Bing** returned around **20%** more webpages that met criteria than **Google** on **Text**.
- **Retailer** was the most common source type for both Google and Bing, returning the most sites classified as meeting the criteria for containing potentially prohibited content, followed by **E-Commerce**. This may be due to the transactional nature of the search queries used.

# The small number of webpages meeting criteria in Image search could be explained by the functionality of the image page indexing

Feature

Image Search

Product Description



PPC Prevalence

43/1152 URLs  
3.7%

This slide builds on the **core research** findings. See methods slides 12-25

The Search Engine Results Page differs depending on what search service product the user is engaging with. According to Google, an **image result** is one that's based on an image that's **embedded on a web page**. An image result is more likely to show for image-seeking queries (1, 2).

The image results page consists of a **series of tiles**, showing an **image thumbnail** for the indexed image, and the **source information** for the web page that's embedding the image (1).

Findings

Websites meeting criteria

Cut of data	#	%
Image/ <b>Total PPC</b>	43/195	22%
Image/ <b>URLs</b>	43/1152	3.7%

Bing/Google

Service	#	%
<b>Google</b>	22/43	51%
<b>Bing</b>	21/43	49%

Source Type

Source Type	#	%
<b>Retailer</b>	32/43	74%
<b>E-Commerce</b>	10/43	23%

Trend Elaboration

- Our research found **fewer websites that met criteria for containing potentially prohibited content** through **image search** as compared to **text search**. This may be due to **multiple results being returned from the same webpage; therefore the proportion of website meeting criteria and websites not meeting criteria could be multiplied**.
- Both **Bing and Google** use 'double-clicking' around image search; meaning that a user could not single-click straight through to the result webpage straight from the search. Rather, they would open an '**expanded search result**', which offered additional functionality such as: **reverse image search** (Google); **rich additional information**, such as price; **related content** not on the core results page.

# Where video results were classified as meeting criteria, they were more likely to be from user-to-user services

**Feature**

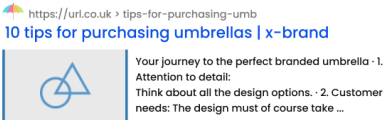
**Video Search**

**PPC Prevalence**

25/1152 URLs  
2%

This slide builds on the **core research** findings. See methods slides 12-25

**Product Description**



According to Google, a video result is a result that's based on a **video that's embedded on a web page**. It's more likely to appear for **video-seeking queries**.

The video results page is more similar to the text-based search results page, with **horizontally stacked results**. Each result consists of a **video thumbnail** (also showing video length), **title link, attribution and upload date** (1, 2).

**Findings**

**Websites meeting criteria**

Cut of data	#	%
Video/ <b>Total PPC</b>	25/195	13%
Video/ <b>URLs</b>	25/1152	2%

**Bing/Google**

Service	#	%
<b>Google</b>	22/25	88%
<b>Bing</b>	3/25	12%

**Source Type**

Source Type	#	%
<b>Social Media</b>	17/25	68%
<b>Retailer</b>	7/25	28%

**Trend Elaboration**

- This research project found fewer websites meeting criteria for containing potentially prohibited content on **video search** compared to **text or image search**. This was most often found on Social Media sites or embedded videos from **Retail sites**.
- Video search returned more **Social Media sites** than any other type of result, and more than other products.
- Bing has its own video site, which is able to play videos from some providers within the Bing domain. This may have contributed to the low score from video search on Bing, if certain providers do not allow their content to be shown.



# The low number of webpages meeting criteria in Shopping may be explained by the paid site indexing functionality

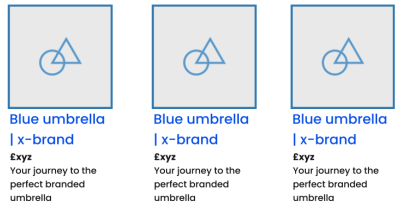
**Feature**

**Product Description**

**Findings**

**Trend Elaboration**

**Shopping Search**



**PPC Prevalence**

16/1152 URLs  
1.4%

This slide builds on the **core research** findings. See methods slides 12-25

Shopping products allow the user to browse products from advertisers and sellers who have **chosen to feature their products** on Google Shopping. Ranking is determined by a combination of **advertiser bids and relevance**. The relative importance of **different user parameters** (user information, google and partner site activity) is determined by **machine learning models**. (1, 2, 3, 4)

The **Shopping results page** is similar to the **Image results page**, consisting of tiles with thumbnails and enriched product details (1).

**Websites meeting criteria**

Cut of data	#	%
Video/ <b>Total PPC</b>	16/195	8%
Video/ <b>V-URLs</b>	16/1152	1.4%

**Bing/Google**

Service	#	%
<b>Google</b>	10/16	62%
<b>Bing</b>	6/16	38%

**Source Type**

Source Type	#	%
<b>Social Media</b>	16/16	100%

- This research project found the **fewest number of sites classified as meeting the criteria for containing potentially prohibited content on Shopping**, as compared to all other products analysed.
- This may be explained by Shopping search product functionality, which indicates that products will **only be surfaced for sponsoring or advertising providers**.
- It may be unlikely that providers trying to sell potentially prohibited items would be paying for **sponsored placements** for them.
- Sites that did score highly under shopping were all **E-commerce sites** which seemed to ship illegally to the UK

# In a limited analysis of "Sponsored Results" within the sample, Google surfaced more websites that met criteria for containing PPC than Bing

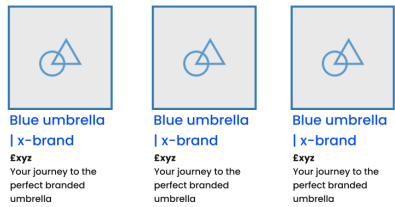
## Feature

## Product Description

## Findings

## Trend Elaboration

### Sponsored Results



### Websites meeting criteria

Cut of data	#	%
PPC/Sponsored	10/95	10%
Spons/URLs	95/252	38%

## PPC Prevalence

10/252 URLs  
38%

We researched **Sponsored Results** across 252 additional URLs to test whether it might surface suggestions to find potentially prohibited items. See full methodology in Slides 12 -25

**Paid search results** appear at the top of the search results page, and they have a green box with the word 'Ad', or have the word 'Sponsored' in bold above them. These are surfaced by providers purchasing **Google Ads/Microsoft Advertising auctions**; only four ads are eligible to show above the search results (1, 2, 3).

The **sponsored results** position on the page is determined by the provider's **Ad Rank**, which is a combination of the bid, auction-time ad quality, the Ad Rank thresholds and the context of the user search (1).

### Bing/Google

Service	#	%
Google	10/10	100%
Bing	0/10	0%

### Product Type

Source Type	#	%
Text	10/10	100%
Image	0/10	0%

- Of the **252 URLs collected, 95 were sponsored results**. However, of these, **10% of these were classified as meeting the criteria for containing potentially prohibited content**. All of the sponsored results that were classified as such were found on **Google Text Search**. Though Bing **returned more sponsored results**, none of them made potentially prohibited items accessible.

# The “related searches” functionality returned similar results to autocomplete, with more suggestions found on Bing

## Feature

### Related Searches

People also ask ⋮

How do you measure umbrella size?

How can you tell a quality umbrella?

Which brand of umbrella is best?

We researched **Related Searches** across 192 additional queries to test whether it might surface suggestions to find potentially prohibited items. See full methodology in Slides 12 - 25

[Related searches](#) appear at the bottom of the search results page. These can consist of **Top Searches** or **Rising Searches**.

**Top Searches:** Terms that are most frequently searched with the user’s term in the same search session, within the chosen category, country or region.

**Rising searches:** Rising searches are terms that were search for with the keyword the user entered which had the most significant growth in volume in the requested time period. Google Trends removes items that may be explicitly sexual, but does not filter controversial topics (1, 2).

## Product Description

## Findings

### Summary Statistics

Response	#	%
<b>Suggestions</b>	<b>120/192</b>	<b>63%</b>
<b>NA</b>	<b>72/192</b>	<b>37%</b>

### Bing/Google

Service	#	%
<b>Google</b>	<b>40/120</b>	<b>40%</b>
<b>Bing</b>	<b>72/120</b>	<b>60%</b>

### Query Type

Query Type	#	%
<b>Root</b>	<b>32/120</b>	<b>27%</b>
<b>Coded</b>	<b>32/120</b>	<b>27%</b>
<b>Extended</b>	<b>56/120</b>	<b>47%</b>

## Trend Elaboration

- **63% of queries** received related search suggestions, and **37% did not**. This is very similar to autocomplete and suggests the underlying algorithms are related.
- Bing delivered **20% more related searches suggestions** on queries than Google for **the same set of queries**. The greatest differences arose across **Cannabis and Stun Guns**.
- **Extended queries** returned many more related searches than other queries, at 47% compared to 27% for Root and Coded. This might suggest that the more detail input by a user, the less likely moderation is to cover it.

# Reflections and Implications

*(Pages 60-63)*

This section looks back over the research project to provide the following:

- Reflections on challenges of the project, which may have implications for future work
- Value of the methodology and findings

# The illicit nature of offers to sell or supply potentially prohibited items pose challenges for future research

Challenge to future or repeated research				
	1	2	3	4
	Ephemerality of content	Obfuscation	Multi-modal evidence	Scam content
Challenge	When returning to conduct quality assurance, many of the results pages had <b>changed or disappeared.</b>	Obfuscation made it difficult to identify some PPC. As a result 50 URLs were not classified as meeting criteria.	Some results used <b>multiple forms of content</b> , i.e. image and text, video and text.	There is a possibility that some high scoring results are <b>scams</b> , according to some specialist indicators.
Outcome	When repeating or replicating this research, care must be taken to <b>rigorously score and document</b> findings.	Engagement with subject matter experts is needed to <b>build on the scoring frameworks.</b>	Research should be conducted into <b>art of the possible</b> for multimodal content moderation.	Engagement with subject matter experts is needed to <b>build on the scoring frameworks.</b>

# Possible avenues for future research have emerged from this project

PUBLIC

## Additional services and features

Additional services could be examined using the **query-based methodology we have established**, such as:

- Generative **AI** services and **chatbots**;
- **User-to-user** services

Additionally the **impact on risk** that other **search service functionalities** may have, such as:

- **User reporting** capabilities.

## Expanding the methodology

The **existing methodology could be broadened** to a broader or different set of hypotheses, such as:

- Expanding to assess **further categories or items** of potentially prohibited items;
- Examining **other, smaller search service platforms**

## Adapting the methodology

To assess risk in light of a **change in user behaviour** the methodology could **change parameters** in the search process, such as:

- Assessing how many potentially prohibited items can be found after **searching for a five minute period**;
- Analysing search results surfaced in **higher pages or tabs**;
- Extending analysis by **following a set number of multiple links or clicks** on sites returned by search queries.

# Appendix

*(Pages 63-69)*

## **Additional resources:**

**A1:** Research methodology

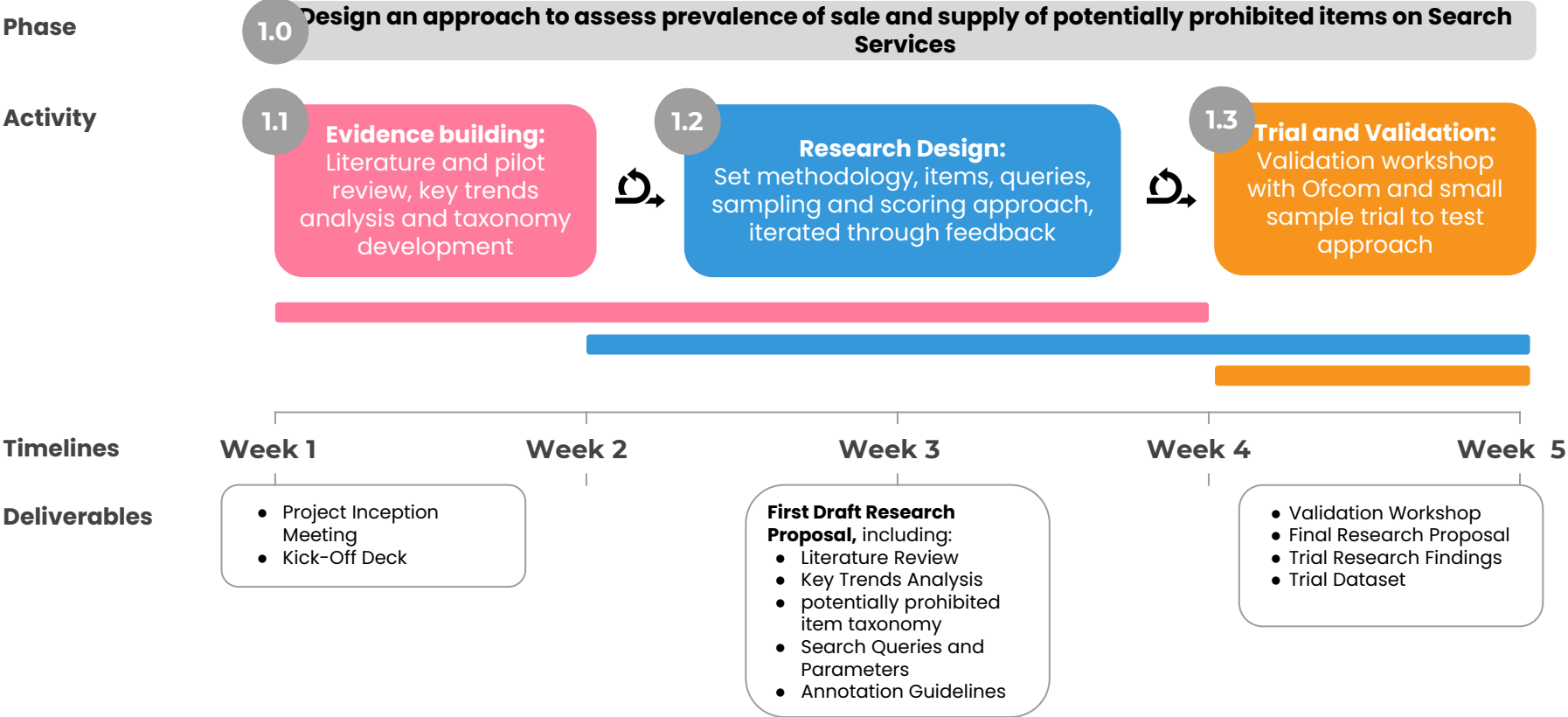
**A2:** Expanded view of research methodology for analysing search engine results pages

**A3, A4:** Detailed trend analysis

**A5:** Validation workshop

**A6:** Glossary of terms

# AI: In Phase 1 we rapidly developed, tested and validated the research design to de-risk fieldwork and maximise insights



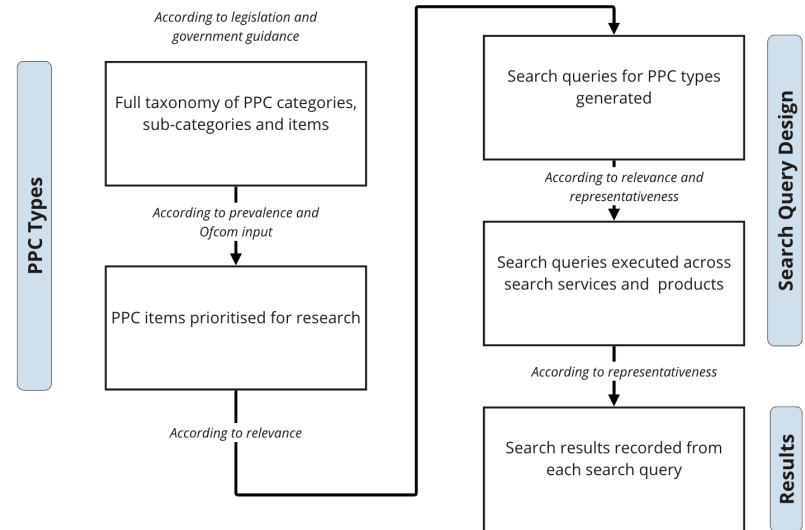


# A2: Search Engine Results Page Methodology

PUBLIC agreed with Ofcom to focus the majority of our research on establishing a robust evidence base in direct response to the core research questions, with a small portion of research dedicated to interrogating some additional hypotheses agreed with Ofcom. The approach to the core research we will carry out can be described as follows:

- 1. Prioritisation of potentially prohibited content Types:** Collaborative downselection with Ofcom of a set number of potentially prohibited content items (i.e., cocaine) from those provided by legislation, to form the basis for search queries.
- 2. Search query design:** Designing a set number of varied search queries for each potentially prohibited content type that will be used as the basis of a typical search. We will ensure that, where appropriate and according to specific potentially prohibited content items, querying may not follow this typical approach.
- 3. Sampling SS:** Executing the search queries across a set of search services and products, and recording a set of the results that are returned.
- 4. Labelling results:** Annotating and scoring results by a set of metrics designed to recognise the presence and likelihood of sale or supply of potentially prohibited content
- 5. Pattern analysis:** Collating search result scores across search services and products and potentially prohibited content types to find patterns across queries.

The sampling process, covering Steps 1-4 above, is illustrated below.



# A3: Search services' functionality has expanded quickly, which may have a mixed impact on availability of potentially prohibited content

## Functionality

## Overview

## Implication for prevalence of PPC

## Overview of UK search services market

### Search Engine Results Page

- Search services **began as a list of hyperlinks** to related web pages but have rapidly evolved.
- The main search engine results page of major search services now **contain multiple widgets that enrich the search experience** and “attempt to provide answers to the query instead of just pointing to websites that might deliver that information”<sup>1</sup>.
- In 2014 Google adjusted its algorithm “so that **illegal websites would appear lower in Google search results**”<sup>2</sup> in response to calls to redirect away from pirated entertainment content being published.

### Additional search products

- In addition to the results page, **Google and Bing offer bespoke tabs** for “Images”, “Video”, “Shopping”, “News”, “Maps” and “Flights”.
- Google also offer “Books” and “Finance”. Bing also offer “Travel” and “Hotels”.
- Yahoo, DuckDuckGo and AOL offer **no additional bespoke products**.

- Widgets that seek to answer queries, rather than provide options **may be less likely to surface potentially prohibited content** as many users are not trying to buy potentially prohibited content and search services do not want to surface it readily.
- **Minimal research** has reviewed the **prevalence of harmful content contained within search engine results page widgets**.
- **Users are more likely to select potentially prohibited content when it appears higher in the results page**<sup>3</sup>. In combination with Google’s efforts to lower these listings on their results page, this **may mean illicit links can be found deeper in search results**.
- Due to the black box nature of search service algorithms, there is **no indication of how shopping results are ordered**<sup>4</sup> so it is unclear how potentially prohibited content availability might be impacted.
- The use of **videos or images may help users find methods of acquiring potentially prohibited content** that is not a classical buyer <> seller. E.g., instructions on how to 3D print a gun<sup>5</sup>.

Search service	UK market share <sup>6</sup>
Google	91.76%
Bing	5.86%
Yahoo	0.86%
DuckDuckGo	0.58%
Ecosia.org	0.29%
Other	0.64%

Source(s): (1) Bruno Oliveira and Carla Teixeira Lopes. 2023. [From 10 Blue Links Pages to Feature-Full Search Engine Results Pages](#), (2) IDATE Consulting on behalf of Ofcom. 2016. [Online Content Study](#), (3) Sivan, Liron & Smith, Michael & Telang, Rahul. 2014. [Do Search Engines Influence Media Piracy? Evidence from a Randomized Field Study](#), (4) [Radd Interactive](#), (5) [Sky News](#). 2022, (6) [Similarweb](#)

# A4: Broader lessons learned from research into illegal online drug supply may be applicable to other forms of PPC

## Where drugs are traded

### Encrypted environments.

There is evidence of drugs being traded in two key encrypted environments:

- The **deep web**<sup>1</sup> (e.g. using The Onion Router)
- **Cryptomarket forums** (e.g. Pretty Good Privacy, Privnote).

### Non-encrypted environments.

There is evidence of **a number of mechanisms in which buyers and sellers of drugs are interacting**<sup>2</sup>, including:

- Private messaging
- Private forums
- Social media platforms
- Spam email

These techniques can often be **used together** with anonymous or ghost accounts.

## How sale of drugs are disguised

### Coded and co-opted language.

- Acronyms and street names **reduce the visibility of the sale of drugs**, for example using the street name for drugs like methamphetamine.
- Existing words are **used out of normal context**, for example 'pizza' instead of cannabis<sup>2</sup>.
- Using completely **different products**, like novel psychoactive substances (NPS) being sold as bath salts and incense sticks amongst others.

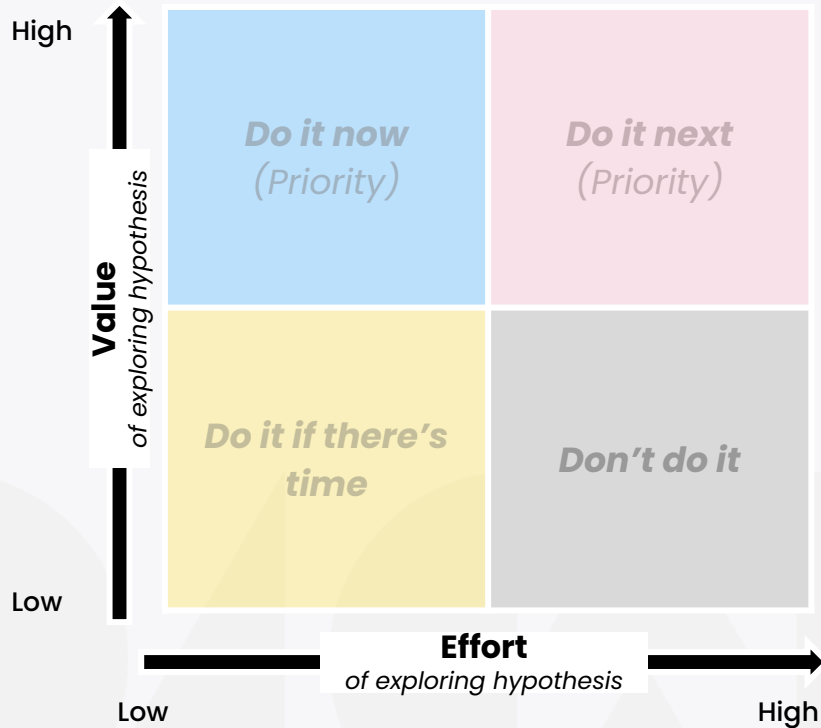
### Changes made to websites.

- **Changing the content on a website**, like advertisements for potentially prohibited content tending to be online for a few days even if the websites may exist for multiple years.
- Websites are often **hidden behind legitimate content** and web addresses that **suggest legal activities**, for example "horse auctions".

## Supply of other forms of PPC

- **No research could be found** about the online supply of **firearms, bladed weapons, or articles used in the commission of fraud**.
- The **venues** where drugs are traded and **methods of disguising sales** may be expected - to some degree - in other areas of PPC.
- There is an indication that, with regards to firearms, the item itself may not be readily available, but **instructions to make your own do exist online**<sup>3</sup>.

# A5: With Ofcom, we prioritised additional hypotheses to understand key platform functionality and the DuckDuckGo search engine



## Priority Areas

Priority areas identified by Ofcom in the validation workshop for immediate investigation included:

- **Platform Functionalities:** Such as Autocomplete, Related Searches and Sponsored Results
- **Alternative Search Services** such as DuckDuckGo

Areas identified as interesting for future investigation included:

- **Algorithmic Assessment and Personalisation**
- **Generative AI** and **Chatbots**

## Takeaways

- Based on this workshop, PUBLIC and Ofcom decided to investigate **Platform Functionalities** and **Alternative Search Services** in an additional research sprint following the primary experiment
- The objective of this sprint was to assess whether these areas had an **impact on results returned**, in comparison to the main research results

## A6: Glossary

Acronym	Definition
GBP	Pound sterling
GHB	Gamma-hydroxybutyrate
MXE	Methoxetamine
OSB	Online Safety Bill
PPC	'Potentially Prohibited Content', or content that contains an apparent offer to sell or supply potentially prohibited items or articles
RQ	Research question
SERP	Search Engine Results Page
SS	Search Services
SEO	Search Engine Optimisation
URL	Uniform Resource Locator, used in this research to refer to a single search result
VPN	Virtual Private Network