Product Scanner AI

White paper | Project vision and proposed capabilities

01 / Project overview

Product Scanner AI is a proposed application that scans barcodes and QR codes on physical products and turns the available information into a detailed, AI-powered product report. It combines information from online sources with text read directly from the product's packaging, helping users understand both what a product is and what is known about it.

Users can request a complete report or focus on a particular question, such as the ingredients in a food product. They can also chat with the AI about the scanned item. The AI draws on the information collected for that product and falls back to a live web search when the requested information is missing from the aggregated sources.

A product profile built from multiple sources

The proposed system brings together product identifiers, structured databases, manufacturer information, category-specific references, safety notices, retail information and packaging photographs. Rather than treating every source as equivalent, the report identifies where important information came from and distinguishes missing, unverified or conflicting information.

Scope and information availability

The aim is to present the maximum useful information supported by the available evidence. Not every product will have every field. Coverage depends on the product, its category, the country or jurisdiction, the sources available and the information visible on the physical package.

The capabilities in this white paper describe the proposed application. Named sources, marketplaces and certification systems illustrate the intended coverage; their inclusion does not establish that an integration or partnership is already operating.

How this document is organized

The following sections describe the user experience, the seven-layer information architecture, the full information catalogue, the role of package scanning, AI consolidation and chat, and two examples of the resulting product report.

Editorial note: This edition reorganizes and clarifies the supplied white paper, including its image-based examples. Overlapping descriptions have been combined while retaining their distinct details.

02 / The user experience

From a scan to an informed product decision

Scan or search for a product

Users can begin by scanning a product's barcode or QR code. The system decodes the available identifiers and uses them to find relevant product information. Users can also search by product name and manufacturing company. Packaging images can support visual recognition of the brand, product and category.

Choose the level of detail

A user can request a fully detailed report or select only the information they want, such as ingredients. This allows the same product profile to support both a quick check and a more thorough review.

Ask follow-up questions

The AI chat lets users ask natural-language questions about the scanned product. Answers use the aggregated product information; when that information does not contain the answer, the proposed system uses a live web-search fallback.

Rate products and explore alternatives

Users will have an option to rate products and view similar products. The app is also intended to assess how a product fits a user's stated health condition or diet. If it does not fit, the app will suggest alternatives. It will also suggest healthier alternatives when the scanned product is assessed as unhealthy. These are proposed personalization features; the source document does not define a scoring method or clinical validation process.

Repeat scans, language support and coupons

- **Free repeat scans:** Scanning the same product again is free.
- **40+ languages:** The app is intended to support more than 40 languages. Product descriptions and labels can be translated while retaining the original-language content.
- **Occasional coupons:** Users may sometimes receive coupons for major marketplaces, such as Walmart. The source describes this as an occasional benefit, not a guaranteed offer for every product or scan.

Understand the evidence behind the result

Important fields carry a source or evidence label. Users should be able to distinguish information supplied by a manufacturer, GS1, a government source, a retailer, a third party or text extracted from packaging. Missing information and disagreements between sources remain part of the report rather than disappearing from the explanation.

03 / Proposed information architecture

Seven layers of product information

The process begins with a scan: decode the barcode or QR code, then identify any available GTIN, EAN, UPC, serial number or lot number. The system uses the available identifiers to match the product against the following information layers.

Layer 1 / Product identity

Barcode and GTIN lookup APIs establish the product's identity and associated identifiers. Where supported, GTIN verification checks validity, identifies the licensed company and helps establish the relationship between the product and that company.

Layer 2 / Product master data

Product information management (PIM) and Global Data Synchronisation Network (GDSN)-grade sources provide structured product records. The source also allows for free alternatives where suitable information is available.

Layer 3 / Category-specific enrichment

Specialized sources add information relevant to the product category, including food, cosmetics, electronics, medicine and automotive products. The information collected depends on the category and available coverage.

Layer 4 / Safety and recall checks

Government and regulatory databases provide recalls, safety alerts and applicable enforcement information. Results need to reflect the relevant country or jurisdiction and, where available, the affected product model or lot.

Layer 5 / Retail and market information

Retail and market sources contribute pricing, availability, offers, listings and reviews where available. Public price observations and merchant or catalogue information supplement the product profile.

Layer 6 / Official manufacturer sources

Manufacturer websites, PDF manuals, product documents and GS1 Digital Link resources contribute official product information, specifications and supporting documentation.

Layer 7 / Physical package scanning

Optical character recognition (OCR) reads text from front, back and side photographs of the package. This includes dedicated extraction of ingredient lists, nutrition panels and allergen warnings, as well as other readable package text.

Output / One consolidated report

The AI combines these layers into a product report, normalizes comparable information, identifies conflicts and marks important fields by source and evidence type: manufacturer, GS1, government, retailer, third-party or OCR-extracted.

04 / Product identity and packaging

The foundation of the product profile

Identity, brand and classification

The application can potentially collect and display:

- Product name, description and product type.
- GTIN, UPC, EAN, the barcode number and other identifiers, together with barcode- or QR-linked resources.
- SKU, model number and catalogue identifier when available.
- Brand name, brand owner and manufacturer.
- Product category, subcategory and classification.
- GTIN validity, licensed-company information and the relationship between the product and the company, where verification is supported.

Product and packaging images

The profile can include official product images, packaging images, image references and front, side and back views where available. Image analysis can also help recognize the product, its brand and its category.

Quantity and physical attributes

Available package and specification information can include:

- Net quantity and its unit; pack size; size or volume; weight; and the number of pieces.
- Dimensions, capacity, materials and other physical attributes.
- Published features, technical attributes and compatibility.

Text and information printed on the package

OCR can capture essentially everything readable in the supplied photographs, including product descriptions and labels, ingredients, nutrition information, directions and other visible text. Specific fields include:

- Batch or lot number and expiration or best-before date, when readable.
- Printed warnings, safety information and package claims.
- Storage, preparation and use instructions.
- Certifications and origin statements printed on the package.

Preserve the distinction between sources

A statement read from a package is information extracted from that package. A matching manufacturer record or regulatory record is a separate source. Keeping these origins visible helps users understand what supports each field, particularly when package text differs from an online record.

All fields remain subject to the source white paper's coverage limits: the product, category, country, source coverage and readability of the physical package determine what can be shown.

05 / Food, cosmetics and personal care

Category-specific composition and usage information

Food: ingredients and allergens

For food products, the app can collect a complete ingredient list when available, together with normalized ingredient information. Allergen information includes declared allergens, allergen warnings and statements such as "may contain."

Food: nutrition and labels

The app can present nutrition facts and available nutrient values, including calories or energy, protein, carbohydrates, fat, saturated fat, sugars, fiber, sodium or salt, and other recorded nutrients. Package scanning can also capture serving size and the printed nutrition table.

Food-label information includes recorded organic, vegan and vegetarian labels, along with other publicly available certification and claim information. Manufacturer food specifications and product images can supplement these fields.

Food: safety and alternatives

Available food safety notices and recall information form part of the product profile. The planned diet and health-condition features can use relevant product information to assess suitability and suggest alternatives, including healthier alternatives when applicable. The original proposal does not specify the rules used to make these assessments.

Cosmetics and personal care: ingredients

For cosmetics and personal-care products, the app can display the International Nomenclature of Cosmetic Ingredients (INCI) ingredient list and normalized ingredient names. Ingredient explanations can describe what individual ingredients are used for and their recorded functions, drawing on public ingredient-reference and regulatory information.

Cosmetics and personal care: use, claims and safety

Other possible fields include cosmetic claims, usage information, product warnings, manufacturer information and cosmetics safety alerts. As with food information, the app should make the source of each important statement visible and indicate when information is unavailable or unverified.

Normalization without losing the original information

The proposed AI normalization process organizes equivalent ingredient names and comparable fields into a consistent structure. Translations can improve accessibility while retaining the original-language label or description. Source provenance remains available so users can understand which record or package text supplied the information.

06 / Technical, medical and specialist products

Specifications, documentation and category identifiers

Technical, electronics and general products

The app can collect technical specifications, structured attributes, features, supported functions and compatibility information. Compatibility may cover devices, models, accessories, parts and systems. Public catalogues, distributor catalogues and reseller information can extend the available product record; automotive products are also included in the proposed category-enrichment scope.

Supporting manufacturer documentation can include user manuals, installation instructions, maintenance instructions, FAQs, troubleshooting, usage guidance and support information. Publicly disclosed warranty information can include duration, conditions and regional terms. Structured information can also be extracted from manufacturer PDFs and datasheets.

Medicines

Where applicable and publicly available, medicine information can include drug identifiers, package information, ingredients, dosage, warnings and official labels. Public drug-labeling and reference information, along with drug recalls and safety notices, can be linked to the product profile.

Medical devices

Potential medical-device fields include the Unique Device Identifier (UDI), manufacturer, model or catalogue number, registration or listing information, and applicable public regulatory information. Medical-device recalls and safety information are also within scope.

Books

Book-specific information can include the International Standard Book Number (ISBN), title, author, publisher, edition and cover.

Music, video games and board games

Music coverage can include release or catalogue information and barcode-linked media metadata. Video-game coverage can include game, title and product metadata. Specialized identifiers and data for board games are also included where relevant and available.

Availability depends on the matched product

These categories describe potential coverage, not a promise that every record contains every field. The usefulness of a technical specification, medicine identifier or specialist catalogue entry depends on whether the available source can be matched to the product being scanned.

07 / Safety, origin, markets and environment

Information beyond the product's basic description

Safety, recalls and regulatory information

The app can show recalls, safety alerts, hazards, remedies and affected lots or models. It can also include applicable government enforcement or recall actions and country- or jurisdiction-specific safety notices. Public conformity, certification and regulatory lookup information may include CE information, declarations of conformity and other compliance documents.

For products with public Safety Data Sheets (SDS), the profile can include hazard, handling and composition information.

Country of origin, manufacturers and companies

Origin information can come from statements printed on the package or supplied by the manufacturer. Publicly disclosed factory, manufacturer and manufacturing-location information can provide additional context. Company information can include identity, jurisdiction, corporate relationships, public filings and corporate-reference information.

Trade information and its limitation

Where applicable, the app can include aggregate import and export statistics by country and product. **Aggregate trade statistics are not shipment-level bill-of-lading intelligence.** They provide broader trade context rather than the detailed shipment history of an individual product.

Prices, offers and market information

Potential fields include public retailer prices, Open Prices observations, retail offers, product listings, availability and merchant or catalogue information. Participating marketplaces and retailers may provide additional product and offer data. Public market-search information and available reviews can also contribute to the profile.

Energy, materials and environmental information

Where the product can be matched, sources may provide energy-label and government energy-registry information, energy class, consumption and model details. Chemical or material registries can provide substance and regulatory information. Environmental datasets and product carbon-footprint datasets may also be included when available.

Certifications

Public certification information may include energy, environmental and textile certifications, as well as FSC, Fairtrade, Rainforest Alliance, OEKO-TEX and other relevant schemes. Printed certification claims and independently available certification records retain their respective source labels.

08 / Why physical package scanning matters

A cereal-box example

Online databases do not necessarily contain all the information printed on the product in a user's hand. Package scanning can recover details that are absent online and connect them to the product identified by the barcode.

The proposed source inventory explicitly includes package barcode scanning and OCR, nutrition-panel OCR, ingredient-list OCR and allergen-warning OCR. Front, back and side photographs can supply additional readable text.

What the barcode can provide

For a cereal box, a barcode lookup can provide the GTIN, EAN or UPC, product identity, brand and available catalogue information.

What the physical package can provide

Readable package photographs can add:

- The exact product name, ingredient list and allergens.
- The nutrition table, net weight and serving size.
- The manufacturer, address and country of origin.
- Warnings and storage instructions.
- Preparation instructions and printed claims.
- Certifications printed on the package.
- The batch or lot number and expiration or best-before date.
- Any other visible text.

How the information comes together

The AI combines barcode data, database records, manufacturer information, package OCR, safety sources, retail information and other matched sources into one report. Information from the package remains distinguishable from information obtained online.

For example, the barcode may identify a cereal product while a photograph supplies the batch number printed on that particular box. Both details belong in the report, but they come from different evidence and should retain that distinction.

Readability and coverage still apply

Package scanning can only extract information that is visible and readable in the supplied images. The original proposal's limits on field availability continue to apply even when multiple information sources are combined.

09 / AI consolidation, evidence and chat

Turning separate records into a coherent explanation

Assemble and normalize the product profile

The AI combines matched records into one consolidated profile. The proposed normalization capabilities include duplicate detection, field normalization, unit conversion or standardization, and matching information across sources. Public-source discovery can locate manufacturer pages, product documents, manuals, specifications and other relevant material.

Detect disagreements and gaps

When two sources provide different information, the system identifies the conflict. It also identifies missing or unverified information, rather than presenting an incomplete field as an established fact.

Show where information came from

Source provenance identifies which source supplied each piece of information. The white paper describes labels including manufacturer, GS1, government, retailer, third-party and OCR-extracted. Its expanded inventory also distinguishes package information, open-database information and secondary-source information. These labels communicate the type of evidence behind a field; the source document does not define a numerical confidence score.

Answer questions and search when necessary

Users can ask natural-language questions about the scanned product. The AI answers using the consolidated data and falls back to a live web search when the requested information is absent from the aggregated sources. Translated descriptions and labels retain their original-language content.

Example user-facing report structure

- **Identity:** Product name, brand, GTIN/EAN/UPC, category and SKU or model.
- **Package:** Size, quantity, labels, warnings and OCR-extracted text.
- **Composition:** Ingredients, allergens and nutrition, where applicable.
- **Specifications:** Technical attributes, dimensions, materials, compatibility and features.
- **Safety:** Recall or safety status and relevant government notices.
- **Origin:** Package- or manufacturer-stated origin and public manufacturer or location information.
- **Retail:** Available prices, offers and availability observations.
- **Documents:** Manuals, warranty, FAQs, support and other public documents.
- **Certifications and environment:** Relevant public certification, energy, chemical and environmental information.
- **Evidence:** The source and confidence or evidence type for important fields.

10 / Illustrative product report

Coca-Cola Original Taste

The following example preserves the fields and sample values in the supplied white paper. It is an illustrative report layout, not a verified product lookup. "X," "Y," "Z" and "XXXXX" are placeholders; fields marked "Not supplied in example" correspond to the original ellipses.

Field	Illustrative value
Product	Coca-Cola Original Taste
Brand	Coca-Cola
GTIN	XXXXX
Package size	330 ml
Category	Soft drink
Manufacturer	Coca-Cola company/entity identified from available sources
Country of sale	X
Country of origin	X
Ingredients	Not supplied in example
Nutrition	Not supplied in example
Allergens	Not supplied in example
Packaging claims	Not supplied in example
Product images	Not supplied in example
Retail prices	X, Y, Z
Available offers	Not supplied in example
Recall status	No matching recall found in checked databases
Manufacturer information	Not supplied in example
Documents	Not supplied in example
Certifications	Not supplied in example
Related products	Not supplied in example
Sources checked	20+
Conflicting information	Not supplied in example
Information not found	Not supplied in example

Interpreting the example

The sample's "20+" source count is illustrative, not a minimum promised for every scan. Likewise, "No matching recall found in checked databases" describes the result of a particular set of checks; it does not establish that every database or jurisdiction has been covered. A real report depends on the matched sources and available product information.

Source basis: Untitled document (2).pdf, nine pages, including the additional product-information inventory, the cereal-box example and the report content embedded as images.