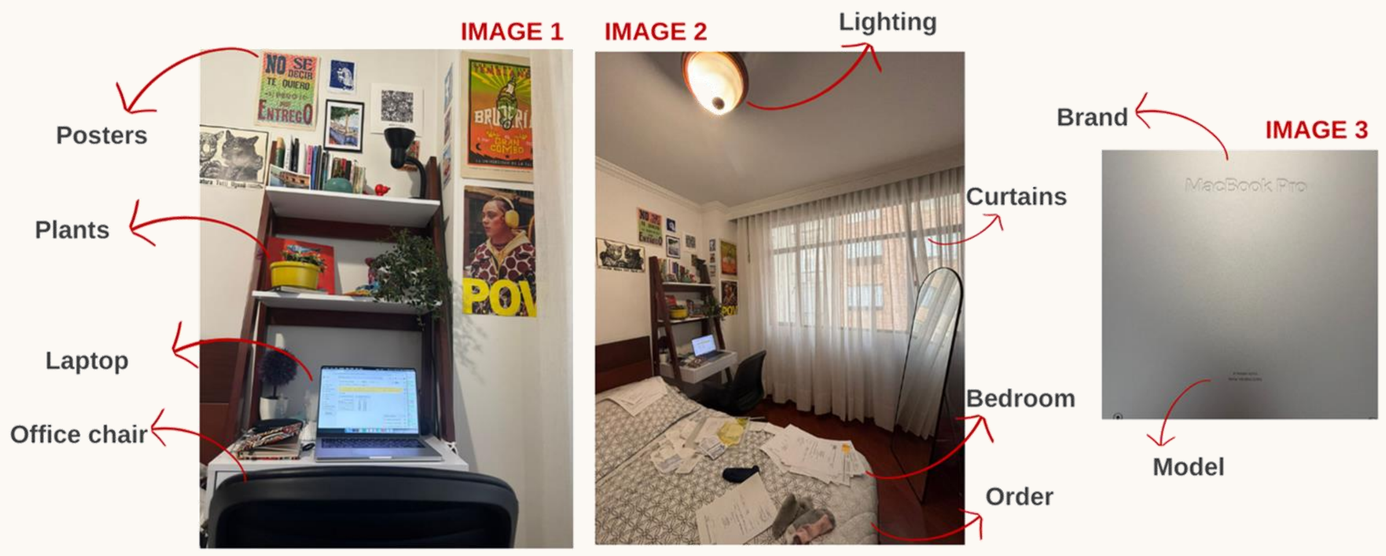
Beyond Algorithms: Why Manual Classification of Visual Data Obtained in Surveys Remains Essential

Mª Paula Acuña-Pardo(RECSM-UPF), Melanie Revilla (RECSM-UPF) & Leyre Padilla (RECSM-UPF)

As **remote work** becomes a stable feature of European employment, assessing remote-working conditions is increasingly important. Conventional survey answers may suffer from measurement errors, linked, for instance, to social desirability bias and respondent burden.

Solution investigated: request **photos of remote workstations to obtain more objective** and detailed information that complements survey measures.



Survey

Opt-in online panel Netquest in Spain

2,326 panelists started + were eligible

Classification objective:

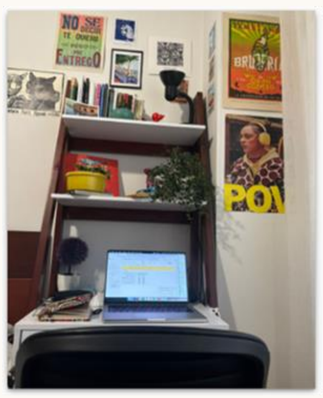
60+ variables from **3 images**

Desks, chairs, equipment, lighting, etc.

Challenge: How to classify?

Vs.

CLOUD VISION API



Electronic Device	98%
Furniture	96%
Office Equipment	93%
Personal Computer	92%
Interior Design	91%
Computer	89%
Computer Hardware	87%
Shelving	87%
Input Device	86%
Peripheral	86%
Display Device	86%
Computer Desk	83%

After scanning the image, it cannot provide most of the information of interest.

CHATGPT

Category	Observation
Desk Type	Compact fixed desk with
Desk Height	Standard
Chair	Office chair with backrest
Laptop	Yes
Desktop	No
Landline	No
Plants	Yes (2)
Decorative Elements	Multiple
Ergonomic Supports	Minimal

GEMINI

Category	Component Details
Furniture Type	**Ladder-style desk** (leaning shelf design) with white shelving and a dark wood frame.
Table Height	**Standard sitting height** (approx. 72-76 cm).
Chair Type	**Ergonomic task chair** with a contoured black backrest.
Working Devices	**Laptop** (Silver MacBook) is the only active device. No land-line, cell-phone, or fixed computer visible.
Peripheral Devices	**Black desk lamp** (positioned on the top shelf).
Ergonomic Elements	**Integrated Backrest** on the chair. No visible footrest, wristrest, or standalone armrests.
Plants	**Three present:** One trailing vine, one potted plant in a yellow bowl, and one purple decorative pompom plant.
Decoration	**Eclectic/Maximalist:** Collection of art posters (e.g., "Brujería"), postcards, a row of books, and small figurines.

Improved performance with detailed prompts, but still limited: For more complex images, the model sometimes generated irrelevant information or ignored specific requested elements. Manual revisions are necessary. Data protection issues.

MANUAL CLASSIFICATION

Still necessary

Procedure

Detailed guidelines

2 classifiers shared the work for Images 1 & 2; 3rd classifier for Image 3

100 cases double-classified → IRR

Disagreements reviewed

Rich information, but challenges

SOME CRUCIAL GUIDELINES

- The classification process is **sequential**.
- Image 2 is used to confirm initial classifying decisions; however, it is classified independently.
- Only **objects located within the room** should be classified.
- **NA:** Used when it does not apply, for instance, if assessing the type of table, but no table is visible.
- **DK:** Used when the object is present, but insufficient visual information prevents reliable classification.
- The presence of an open laptop implies that a keyboard, touchpad, and camera exist and are available for use, even if we do not explicitly see them in the photo.

MAIN CHALLENGES OF MANUAL CLASSIFICATION

Issue	Procedure/ Rule applied
Defining the criteria for considering that an image is classifiable and in line	If the photo shows a space that could likely be a remote workstation, consider it in line. Photos are classified if the main elements can be identified, even if blurred.
Defining the variables to be classified and their labels	Once a first set of variables and labels was defined, classifiers trained with photos taken by the team. This was used to revise the set of variables/labels before the final classification.
Personally Identifiable Information (PII) in photos	All images were reviewed by the ethics advisor before classifying + classifiers were instructed to blur any kind of remaining PII.
Differences across classifiers	IRR was computed + differences were revised by the 3rd researcher. Guidelines were improved to reduce inconsistencies for the remaining classification process. When counts still differ, the highest number selected (greater detection confidence).

Key Findings & Conclusions

IRR

	% Agreement*	Cohen's-Kappa**
MIN	82.4	-0.01
MAX	100	1
AVG	96.4	0.8
MEDIAN	97.2	0.88

*73.3% of variables have a % Agr. >95

**70% of variables have a Kappa ≥.81

Main classifiers disagreements: DECO and No. of WINDOWS

- Visual data adds information not obtainable through conventional questions, but **ethical screening** was crucial since 146 photos were detected with PII, and **classifying** the information is challenging.
- Although AI tools offer important improvements, they still require human supervision to ensure accurate and reliable classification. Moreover, many AI tools raise ethical and data protection concerns.
- Manual classification captures **context** and usability of objects.
- The process is rigorous but time- and labor-intensive. **12 images per hour approx.**

Scan for details