CERTLABZ CERTIFICATION BOARD

Certified Privacy Engineering Specialist (CPES)

Ten Sample Questions

These ten questions are written at the difficulty of the live exam. Six carry a figure, a data table, a configuration or a code listing; four are plain scenario questions, as on the live paper. Every one is a scenario a working privacy engineer would meet, and none of them appears in the live question bank.

The live exam is multistage adaptive: how you score on one stage selects the difficulty of the next, so no two candidates sit the same paper. It is a performance-based exam: each form can carry up to 5 to 12 performance-based questions alongside the multiple-choice questions, which means you build, sequence or complete something rather than only choosing from a list. Answers and full explanations follow each question here; the live exam does not disclose item-level answers.

Question 1 — Domain 3: Data Governance, Mapping and Minimisation — DATA

Emma Wilson reviews a data map for a signup flow. The fields collected and their stated purpose are below.

Field collected	Stated purpose	Retention
Email	Account login and service email	Life of account
Date of birth	Age-gate check (18+) at signup	Life of account
Full home address	None recorded	Life of account
Marketing opt-in flag	Consent record for newsletters	Until withdrawn

Table 1.1 Fields, purposes and retention for the signup flow.

Which finding is the most important privacy issue here?

A. The marketing opt-in flag should not be stored at all, because consent records are unnecessary once a user has signed up.

B. The email is retained too long and should be deleted a fixed thirty days after each individual login event occurs.

C. The home address is collected with no purpose, and the date of birth could be reduced to an over-18 boolean at signup.

D The date of birth must be kept for the full life of the account so the age-gate can be re-checked on every login attempt.

Correct answer: C
Two minimisation failures stand out: the home address is collected with no recorded purpose (collect nothing you cannot justify), and the date of birth is more than the age-gate needs, which only requires an over-18 result that can be stored as a boolean. Consent records are legitimately retained, and the email has a valid ongoing purpose.

Question 2

Domain 5: Privacy in Systems and Engineering

CONFIG

James Carter reviews how a feature records marketing consent.

```
consent = {
  marketing: true,          // set when the user submitted the form
  // no record of when, which notice version, or how to withdraw
}
# withdrawal: user emails support, who edits the flag manually
```

Listing 2.1 The consent record as implemented.

Why is this consent implementation not defensible, and what should change?

A It is fine as written, because a single boolean flag set from the form is all that valid marketing consent actually requires.

B It records no purpose, timestamp, notice version or self-service withdrawal, so it cannot evidence valid, freely revocable consent.

C The only problem is that the flag defaults to true, and flipping the default to false would make the whole record defensible.

D The only fix needed is to encrypt the boolean flag at rest so that the stored consent value cannot be read by an attacker.

Correct answer: B
Valid consent must be specific, informed, evidenced and as easy to withdraw as to give. A lone boolean with no purpose, timestamp or notice version cannot prove valid consent, and manual withdrawal by email is not the easy, self-service revocation required. Encryption and the default are minor next to the missing evidence and withdrawal path.

Question 3

Domain 4: Privacy-Enhancing Technologies

DATA

Olivia Davis must publish a dataset for external research. Four preparations are proposed.

Preparation	Property
1. Remove names, keep DOB, ZIP and sex	Quasi-identifiers remain
2. Unsalted hash of the email as the key	Deterministic, guessable input
3. Generalise quasi-identifiers to k-anonymity with l-diversity	Group size and sensitive-value diversity enforced
4. Keep raw data, encrypt the file	Recipient decrypts to raw data

Table 3.1 Proposed preparations for external release.

Which preparation is the soundest basis for an external release, and why?

A. Preparation 1, because removing the direct names is enough to make the dataset anonymous for research use.

B. Preparation 2, because hashing the email key reliably prevents anyone from re-identifying the individuals.

C. Preparation 4, because encrypting the file means the raw personal data is fully protected once it is shared.

D. Preparation 3, because k-anonymity with l-diversity limits singling-out and attribute disclosure in the released data.

Correct answer: D
Generalising quasi-identifiers to k-anonymity with l-diversity limits both singling-out and attribute disclosure in the released data. Removing names alone leaves re-identifiable quasi-identifiers, an unsalted hash is a guessable pseudonym rather than anonymisation, and encrypting the file simply hands the recipient the raw personal data once decrypted.

Question 4

Domain 5: Privacy in Systems and Engineering

DATA

Liam Anderson designs deletion for a subject who exercises the right to erasure. Their data sits in the places below.

Location	Proposed handling
Primary database	Delete the record immediately
Search index	(to decide)
Nightly backups	(to decide)
Third-party CRM	(to decide)

Table 4.1 Where the data lives and what to do.

Which set of handling decisions correctly completes the erasure?

A. Search index: leave it, indexes are derived; backups: never touch; CRM: assume the vendor deletes it automatically over time.

B. Search index: re-index to drop the record; backups: purge or expire within a documented window; CRM: trigger contractual deletion.

C Search index: encrypt the entry; backups: restore the record to preserve it; CRM: export the data before any deletion happens.

D Search index: ignore it; backups: keep for disaster recovery only; CRM: out of scope because it is a separate company entirely.

Correct answer: B
Erasure must reach every copy. The derived search index must be updated to drop the record, backups must be purged or expired within a documented window with restore suppression, and the processor CRM must be told to delete under the contract. Leaving indexes, keeping backups untouched or treating the processor as out of scope all leave the person's data in place.

Question 5

Domain 2: Privacy by Design and Threat Modeling

CONFIG

Ava Thompson threat-models an analytics design that attaches a persistent unique identifier to every event.

```
event = {
  advertising_id: "a7f3...persistent, never rotated",
  timestamp, page, referrer, precise_geolocation
}
# sent to a third-party analytics SDK on every page
```

Listing 5.1 The event payload sent to the analytics SDK.

Which LINDDUN threats does this design most raise, and what mitigates them?

A Availability and denial of service; the mitigation is to add more analytics servers and cache the events at the edge.

B Non-repudiation only; the mitigation is to remove any digital signatures from the analytics events before they are sent.

C Linkability and identifiability; mitigate with rotating or scoped pseudonyms, dropping precise geolocation and minimising fields.

D None worth acting on; a persistent advertising identifier attached to every event is a standard and privacy-safe analytics design.

Correct answer: C
A persistent unique identifier links all of a person's activity, and precise geolocation plus referrer make them identifiable, so linkability and identifiability are the threats. Rotating or scoping the identifier, dropping precise location and minimising the fields sent to the third party directly reduce both. The other options misread the threats entirely.

Question 6

Domain 6: Privacy Programme, Assessments and Breach

DATA

A breach is confirmed. The exposed data and context are below.

Exposed	Detail
Email addresses	Plaintext, 40,000 EU residents
Passwords	Hashed with SHA-256, unsalted
Discovery	Confirmed on a Monday morning
Publication	No evidence of public posting yet

Table 6.1 The breach at a glance.

What does the GDPR require here?

A. Nothing yet, because there is no evidence the exposed data has actually been posted publicly anywhere online so far.

B. Notify the authority within thirty days, and only tell the affected individuals if the data later appears for sale online.

C. Notify the authority without undue delay and within 72 hours where feasible, and notify individuals given the high risk to them.

D. Notify only the affected individuals, since regulator notification is optional whenever passwords were stored in a hashed form.

Correct answer: C
Unsalted hashes are weak and the emails identify 40,000 people, so this is a high risk to individuals. The GDPR requires notifying the supervisory authority without undue delay and within 72 hours where feasible, and notifying the affected individuals because the risk to them is high. Waiting for publication, a 30-day window, or skipping the regulator are all wrong.

Question 7 Domain 1: Privacy Foundations, Principles and Law

A product team wants to reuse customer support transcripts, originally collected to resolve tickets, to train a new marketing model. What is the core privacy issue?

A. There is no issue at all, because the company already holds the transcripts and may use its own data for any purpose it likes.

B. The only issue is storage cost, which is resolved by moving the transcripts to a cheaper archive before the model is trained.

C. The only issue is security, so encrypting the transcripts at rest fully addresses the concern before they are used for training.

D. It is a new, likely incompatible purpose that needs its own lawful basis and transparency, not just technical access to the data.

Correct answer: D
Purpose limitation applies within a single company. Support data collected to resolve tickets cannot simply be repurposed to train a marketing model; that is a new and likely incompatible purpose requiring its own lawful basis, transparency and possibly consent. Ownership, storage and encryption do not resolve the purpose problem.

Question 8 Domain 4: Privacy-Enhancing Technologies

A team hashes email addresses with unsalted SHA-256 and describes the resulting dataset as "anonymised". Why is that description wrong?

A. It is not wrong; a cryptographic hash of an email address reliably anonymises it and removes it from the scope of privacy law.

B. The hash is deterministic over a guessable input, so the same person maps to the same value and can be re-identified; it is a pseudonym.

C. It is wrong only because SHA-256 is outdated, and switching to a newer hashing algorithm would make the dataset genuinely anonymous.

D. It is wrong only because the hash was not encrypted afterwards, and encrypting the hashed values would complete the anonymisation.

Correct answer: B
A hash of a known-format identifier is a pseudonym, not anonymisation: it is deterministic and the input space is small enough to brute-force or match, so individuals remain re-identifiable and the data is still personal. Changing the algorithm or encrypting the output does not change that; only breaking the link irreversibly would.

Question 9 Domain 2: Privacy by Design and Threat Modeling

A recommendation feature could either send each user's full browsing history to the server or run inference on the device. Which is the privacy-preferable design, and why?

A. Server-side collection, because holding the raw browsing history centrally is always better for building accurate recommendations.

B. The two designs are equivalent for privacy, so the choice should be made purely on which one is cheaper and faster to build.

C. On-device inference, because it avoids centrally collecting the raw browsing history, applying minimisation and cutting exposure.

D. Server-side collection, because central data is easier to secure and encryption removes any privacy difference between the two.

Correct answer: C
Not collecting the raw data centrally is the strongest privacy control available. On-device inference delivers the feature while avoiding a central store of browsing histories, which reduces breach scope, linkability and governance burden. The other options ignore that avoided collection beats after-the-fact protection.

Question 10 Domain 6: Privacy Programme, Assessments and Breach

A DPIA concludes that a planned processing activity carries a high residual risk that the proposed mitigations cannot sufficiently reduce. What does the GDPR require next?

A. Start the processing and monitor it closely, on the basis that a completed DPIA is itself sufficient to proceed with high risk.

B. Consult the supervisory authority before starting, because unmitigated high residual risk triggers prior consultation under the GDPR.

C. Delete the DPIA and proceed quietly, since a documented finding of high residual risk would otherwise create unnecessary liability.

D. Reduce the number of affected records below one million and then begin the processing without any further external consultation.

Correct answer: B

When a DPIA shows a high residual risk that mitigations cannot bring down, the GDPR requires prior consultation with the supervisory authority before the processing begins. Proceeding anyway, hiding the assessment, or simply shrinking the record count are not lawful responses to unmitigated high risk.

Published by the CertLabz Certification Board, a division of CertLabz LLC. Companion to objectives document CPES-001. https://certlabz.com/certifications/certified-privacy-engineering-specialist.html