Submission and Formatting Instructions for International Conference on Machine Learning (ICML 2025)

Firstname1 Lastname1 [*] [1] Firstname2 Lastname2 [*] [1] [2] Firstname3 Lastname3 [2] Firstname4 Lastname4 [3]
Firstname5 Lastname5 [1] Firstname6 Lastname6 [3] [1] [2] Firstname7 Lastname7 [2] Firstname8 Lastname8 [3]
Firstname8 Lastname8 [1] [2]

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1. Electronic Submission

Submission to ICML 2025 will be entirely electronic, via a web site (not email). Information about the submission process and LaTeX templates are available on the conference web site at:

http://icml.cc/

The guidelines below will be enforced for initial submissions and camera-ready copies. Here is a brief summary:

- Submissions must be in PDF.

- If your paper has appendices, submit the appendix together with the main body and the references as a single file. Reviewers will not look for appendices as a separate PDF file. So if you submit such an extra file, reviewers will very likely miss it.

- Page limit: The main body of the paper has to be fitted to 8 pages, excluding references and appendices; the space for the latter two is not limited in pages, but the total file size may not exceed 10MB. For the final version of the paper, authors can add one extra page to the main body.

- Do not include author information or acknowledgements in your initial submission.

- Your paper should be in 10 point Times font.

- Make sure your PDF file only uses Type-1 fonts.

- Place figure captions under the figure (and omit titles from inside the graphic file itself). Place table captions over the table.

- References must include page numbers whenever possible and be as complete as possible. Place multiple citations in chronological order.

- Do not alter the style template; in particular, do not compress the paper format by reducing the vertical spaces.

- Keep your abstract brief and self-contained, one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase. The title should have content words capitalized.

1.1. Deep Q-Network with Experience Replay

The following pseudocode gives a concise implementation flow for a Deep Q-Network (DQN) using an experience replay buffer and a target network. It is written in the standard algorithm / algorithmic environments and suitable as a high-level recipe for experiments.

Notes. Typical practical details: use RMSProp/Adam for optimization; clip rewards or normalize observations when appropriate; use a separate periodically-updated target network to stabilise learning; tune replay capacity N and batch size B; optionally use prioritized replay or Double DQN extensions.

---

[*]Equal contribution [1]Department of XXX, University of YYY, Location, Country [2]Company Name, Location, Country [3]School of ZZZ, Institute of WWW, Location, Country. Correspondence to: Firstname1 Lastname1 <first1.last1@xxx.edu>, Firstname2 Lastname2 <first2.last2@www.uk>.

Algorithm 1 DQN with Experience Replay

1: Input: environment Env, replay capacity N, batch size B, discount γ, target update period C, learning rate α, exploration schedule ϵ_t
2: Initialize replay buffer $\mathcal{D} \leftarrow \{\}$ of capacity N
3: Initialize online Q-network $Q(s, a; \theta)$ with random weights θ
4: Initialize target network weights $\theta^- \leftarrow \theta$
5: for episode = 1 to M do
6:    $s \leftarrow$ Env.reset()
7:    for t = 1 to T do
8:       With probability ϵ_t select a random action a, otherwise $a \leftarrow \arg\max_{a'} Q(s, a'; \theta)$
9:       Execute a, observe reward r and next state s'
10:      Store transition (s, a, r, s') in $\mathcal{D}$ (drop oldest if full)
11:      if $|\mathcal{D}| \geq B$ then
12:         Sample random batch $\{(s_j, a_j, r_j, s'_j)\}_{j=1}^{B}$ from $\mathcal{D}$
13:         For each sample compute target:

$$y_j = \begin{cases} r_j & \text{if } s'_j \text{ is terminal} \\ r_j + \gamma \max_{a'} Q(s'_j, a'; \theta^-) & \text{otherwise} \end{cases}$$

14:         Perform a gradient step on loss $\mathcal{L}(\theta) = \frac{1}{B} \sum_j (y_j - Q(s_j, a_j; \theta))^2$
15:      end if
16:      Every C steps: $\theta^- \leftarrow \theta$
17:      $s \leftarrow s'$
18:   end for
19: end for

1.2. Double Deep Q-Network (Double DQN)

Double DQN reduces the overestimation bias of the original DQN by decoupling the action selection and action evaluation between the online network and the target network. The pseudocode below shows this modification to the target computation while keeping the overall training loop and replay buffer logic identical to standard DQN.

Notes on Double DQN. Compared to vanilla DQN, Double DQN uses the online network θ to select the argmax action and the target network θ^- to evaluate its value, which empirically reduces overestimation and often improves stability. Other recommended improvements (e.g., prioritized replay, dueling architecture, n-step returns, Huber loss) are compatible with this change.

1.3. Dueling Deep Q-Network (Dueling DQN)

Dueling DQN separates the representation of state-value and advantage for each action, which helps the agent learn which states are (or are not) valuable without having to learn the effect of each action for every state. The architecture splits the final layers of the Q-network into two streams that estimate a scalar state-value $V(s)$ and an advantage vector $A(s, a)$; these are combined to produce Q-values via $Q(s, a) = V(s) + A(s, a) - \frac{1}{|\mathcal{A}|} \sum_{a'} A(s, a')$.

The pseudocode below shows a dueling variant that uses experience replay and a target network; it otherwise follows the same training loop as the DQN family (optionally usable together with Double DQN selection/evaluation).

Notes on Dueling DQN. The dueling architecture often speeds learning of state values and leads to more stable policies, particularly in environments where many actions have similar effects. Ensure advantage normalization (subtracting the mean advantage) when reconstructing Q-values to guarantee identifiability of V and A. Dueling is orthogonal to other improvements (Double DQN, prioritized replay, n-step returns) and can be combined with them.

1.4. Submitting Papers

Anonymous Submission: ICML uses double-blind review: no identifying author information may appear on the title page or in the paper itself. Section 2.3 gives further details.

Authors must provide their manuscripts in PDF

Algorithm 2 Double DQN with Experience Replay

1: Input: environment Env, replay capacity N, batch size B, discount γ, target update period C, learning rate α, exploration schedule ϵ_t
2: Initialize replay buffer $\mathcal{D} \leftarrow \{\}$ of capacity N
3: Initialize online Q-network $Q(s, a; \theta)$ with random weights θ
4: Initialize target network weights $\theta^- \leftarrow \theta$
5: for episode = 1 to M do
6:    $s \leftarrow$ Env.reset()
7:    for t = 1 to T do
8:       With probability ϵ_t select a random action a, otherwise $a \leftarrow \arg\max_{a'} Q(s, a'; \theta)$
9:       Execute a, observe reward r and next state s'
10:      Store transition (s, a, r, s') in $\mathcal{D}$ (drop oldest if full)
11:      if $|\mathcal{D}| \geq B$ then
12:        Sample random batch $\{(s_j, a_j, r_j, s'_j)\}_{j=1}^{B}$ from $\mathcal{D}$
13:        For each sample compute Double DQN target:

$$y_j = \begin{cases} r_j & \text{if } s'_j \text{ is terminal} \\ r_j + \gamma \, Q\big(s'_j, \arg\max_a Q(s'_j, a; \theta); \theta^-\big) & \text{otherwise} \end{cases}$$

14:        Perform a gradient step on loss $\mathcal{L}(\theta) = \frac{1}{B} \sum_j (y_j - Q(s_j, a_j; \theta))^2$
15:      end if
16:      Every C steps: $\theta^- \leftarrow \theta$
17:      $s \leftarrow s'$
18:    end for
19: end for

---

format. Furthermore, please make sure that files contain only embedded Type-1 fonts (e.g., using the program pdffonts in linux or using File/DocumentProperties/Fonts in Acrobat). Other fonts (like Type-3) might come from graphics files imported into the document.

Authors using Word must convert their document to PDF. Most of the latest versions of Word have the facility to do this automatically. Submissions will not be accepted in Word format or any format other than PDF. Really. We're not joking. Don't send Word.

Those who use LaTeX should avoid including Type-3 fonts. Those using latex and dvips may need the following two commands:

dvips -Ppdf -tletter -G0 -o paper.ps paper.dvi
ps2pdf paper.ps

It is a zero following the "-G", which tells dvips to use the config.pdf file. Newer TeX distributions don't always need this option.

Using pdflatex rather than latex, often gives better results. This program avoids the Type-3 font problem, and supports more advanced features in the microtype package.

Graphics files should be a reasonable size, and in-

cluded from an appropriate format. Use vector formats (.eps/.pdf) for plots, lossless bitmap formats (.png) for raster graphics with sharp lines, and jpeg for photo-like images.

The style file uses the hyperref package to make clickable links in documents. If this causes problems for you, add nohyperref as one of the options to the icml2025 usepackage statement.

1.5. Submitting Final Camera-Ready Copy

The final versions of papers accepted for publication should follow the same format and naming convention as initial submissions, except that author information (names and affiliations) should be given. See Section 2.3.2 for formatting instructions.

The footnote, "Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute." must be modified to "Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada, PMLR 267, 2025. Copyright 2025 by the author(s)."

For those using the LaTeX style file, this change (and others) is handled automatically by simply changing

---

Algorithm 3 Dueling DQN with Experience Replay

1: Input: environment Env, replay capacity N, batch size B, discount γ, target update period C, learning rate α, exploration schedule ϵ_t
2: Initialize replay buffer $\mathcal{D} \leftarrow \{\}$ of capacity N
3: Initialize online dueling Q-network: shared trunk; value head $V(s; \theta)$; advantage head $A(s, a; \theta)$; combine to $Q(s, a; \theta)$ via advantage normalization
4: Initialize target network weights $\theta^- \leftarrow \theta$
5: for episode = 1 to M do
6:    $s \leftarrow$ Env.reset()
7:    for t = 1 to T do
8:       With probability ϵ_t select a random action a, otherwise $a \leftarrow \arg\max_{a'} Q(s, a'; \theta)$
9:       Execute a, observe reward r and next state s'
10:      Store transition (s, a, r, s') in $\mathcal{D}$ (drop oldest if full)
11:      if $|\mathcal{D}| \geq B$ then
12:        Sample random batch $\{(s_j, a_j, r_j, s'_j)\}_{j=1}^{B}$ from $\mathcal{D}$
13:        For each sample compute target y_j (use either DQN or Double DQN style target):

$$y_j = \begin{cases} r_j & \text{if } s'_j \text{ is terminal} \\ r_j + \gamma \max_{a'} Q(s'_j, a'; \theta^-) & \text{otherwise} \end{cases}$$

14:        Note: when combining with Double DQN, replace the max evaluation by

$$r_j + \gamma\, Q\big(s'_j, \arg\max_a Q(s'_j, a; \theta); \theta^-\big)$$

15:        Perform a gradient step on loss $\mathcal{L}(\theta) = \frac{1}{B}\sum_j (y_j - Q(s_j, a_j; \theta))^2$; gradients backpropagate through both value and advantage heads
16:      end if
17:      Every C steps: $\theta^- \leftarrow \theta$
18:      $s \leftarrow s'$
19:    end for
20: end for

`\usepackage{icml2025}` to

    `\usepackage[accepted]{icml2025}`

Authors using Word must edit the footnote on the first page of the document themselves.

Camera-ready copies should have the title of the paper as running head on each page except the first one. The running title consists of a single line centered above a horizontal rule which is 1 point thick. The running head should be centered, bold and in 9 point type. The rule should be 10 points above the main text. For those using the LaTeX style file, the original title is automatically set as running head using the fancyhdr package which is included in the ICML 2025 style file package. In case that the original title exceeds the size restrictions, a shorter form can be supplied by using

`\icmltitlerunning{...}`

just before `\begin{document}`. Authors using Word must edit the header of the document themselves.

2. Format of the Paper

All submissions must follow the specified format.

2.1. Dimensions

The text of the paper should be formatted in two columns, with an overall width of 6.75 inches, height of 9.0 inches, and 0.25 inches between the columns. The left margin should be 0.75 inches and the top margin 1.0 inch (2.54 cm). The right and bottom margins will depend on whether you print on US letter or A4 paper, but all final versions must be produced for US letter size. Do not write anything on the margins.

The paper body should be set in 10 point type with a vertical spacing of 11 points. Please use Times typeface throughout the text.

2.2. Title

The paper title should be set in 14 point bold type and centered between two horizontal rules that are 1 point thick, with 1.0 inch between the top rule and the top edge of the page. Capitalize the first letter of content words and put the rest of the title in lower case.

2.3. Author Information for Submission

ICML uses double-blind review, so author information must not appear. If you are using LaTeX and the icml2025.sty file, use \icmlauthor{...} to specify authors and \icmlaffiliation{...} to specify affiliations. (Read the TeX code used to produce this document for an example usage.) The author information will not be printed unless accepted is passed as an argument to the style file. Submissions that include the author information will not be reviewed.

2.3.1. Self-Citations

If you are citing published papers for which you are an author, refer to yourself in the third person. In particular, do not use phrases that reveal your identity (e.g., "in previous work (?), we have shown ...").

Do not anonymize citations in the reference section. The only exception are manuscripts that are not yet published (e.g., under submission). If you choose to refer to such unpublished manuscripts (?), anonymized copies have to be submitted as Supplementary Material via OpenReview. However, keep in mind that an ICML paper should be self contained and should contain sufficient detail for the reviewers to evaluate the work. In particular, reviewers are not required to look at the Supplementary Material when writing their review (they are not required to look at more than the first 8 pages of the submitted document).

2.3.2. Camera-Ready Author Information

If a paper is accepted, a final camera-ready copy must be prepared. For camera-ready papers, author information should start 0.3 inches below the bottom rule surrounding the title. The authors' names should appear in 10 point bold type, in a row, separated by white space, and centered. Author names should not be broken across lines. Unbolded superscripted numbers, starting 1, should be used to refer to affiliations.

Affiliations should be numbered in the order of appearance. A single footnote block of text should be used to list all the affiliations. (Academic affiliations should list Department, University, City, State/Region, Country. Similarly for industrial affiliations.)

Each distinct affiliations should be listed once. If an author has multiple affiliations, multiple superscripts should be placed after the name, separated by thin spaces. If the authors would like to highlight equal contribution by multiple first authors, those authors should have an asterisk placed after their name in superscript, and the term "*Equal contribution" should be placed in the footnote block ahead of the list of affiliations. A list of corresponding authors and their emails (in the format Full Name <email@domain.com>) can follow the list of affiliations. Ideally only one or two names should be listed.

A sample file with author names is included in the ICML2025 style file package. Turn on the [accepted] option to the stylefile to see the names rendered. All of the guidelines above are implemented by the LaTeX style file.

2.4. Abstract

The paper abstract should begin in the left column, 0.4 inches below the final address. The heading 'Abstract' should be centered, bold, and in 11 point type. The abstract body should use 10 point type, with a vertical spacing of 11 points, and should be indented 0.25 inches more than normal on left-hand and right-hand margins. Insert 0.4 inches of blank space after the body. Keep your abstract brief and self-contained, limiting it to one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase.

2.5. Partitioning the Text

You should organize your paper into sections and paragraphs to help readers place a structure on the material and understand its contributions.

2.5.1. Sections and Subsections

Section headings should be numbered, flush left, and set in 11 pt bold type with the content words capitalized. Leave 0.25 inches of space before the heading and 0.15 inches after the heading.

Similarly, subsection headings should be numbered, flush left, and set in 10 pt bold type with the content words capitalized. Leave 0.2 inches of space before the heading and 0.13 inches afterward.

Finally, subsubsection headings should be numbered, flush left, and set in 10 pt small caps with the content words capitalized. Leave 0.18 inches of space before the heading and 0.1 inches after the heading.

Please use no more than three levels of headings.

2.5.2. Paragraphs and Footnotes

Within each section or subsection, you should further partition the paper into paragraphs. Do not indent the first line of a given paragraph, but insert a blank line between succeeding ones.

You can use footnotes[1] to provide readers with additional information about a topic without interrupting the flow of the paper. Indicate footnotes with a number in the text where the point is most relevant. Place the footnote in 9 point type at the bottom of the column in which it appears. Precede the first footnote in a column with a horizontal rule of 0.8 inches.[2]

Algorithm 4 Bubble Sort

Input: data x_i, size m
repeat
   Initialize $noChange = true$.
   for $i = 1$ to $m - 1$ do
     if $x_i > x_{i+1}$ then
       Swap x_i and x_{i+1}
       $noChange = false$
     end if
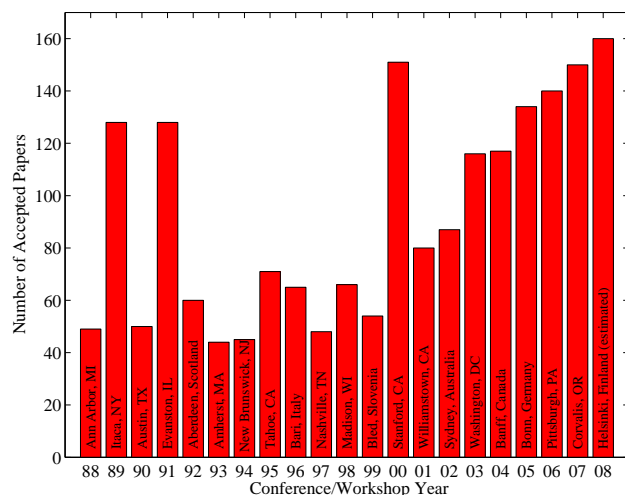   end for
until $noChange$ is $true$

Figure 1. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead estimated.

2.6. Figures

You may want to include figures in the paper to illustrate your approach and results. Such artwork should be centered, legible, and separated from the text. Lines should be dark and at least 0.5 points thick for purposes of reproduction, and text should not appear on a gray background.

Label all distinct components of each figure. If the figure takes the form of a graph, then give a name for each axis and include a legend that briefly describes

each curve. Do not include a title inside the figure; instead, the caption should serve this function.

Number figures sequentially, placing the figure number and caption after the graphics, with at least 0.1 inches of space before the caption and 0.1 inches after it, as in Figure 1. The figure caption should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left. You may float figures to the top or bottom of a column, and you may set wide figures across both columns (use the environment figure* in LaTeX). Always place two-column figures at the top or bottom of the page.

2.7. Algorithms

If you are using LaTeX, please use the "algorithm" and "algorithmic" environments to format pseudocode. These require the corresponding stylefiles, algorithm.sty and algorithmic.sty, which are supplied with this package. Algorithm 4 shows an example.

2.8. Tables

You may also want to include tables that summarize material. Like figures, these should be centered, legible, and numbered consecutively. However, place the title above the table with at least 0.1 inches of space before the title and the same after it, as in Table 1. The table title should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left.

Tables contain textual material, whereas figures contain graphical material. Specify the contents of each row and column in the table's topmost row. Again, you may float tables to a column's top or bottom, and set wide tables across both columns. Place two-column tables at the top or bottom of the page.

---

[1] Footnotes should be complete sentences.

[2] Multiple footnotes can appear in each column, in the same order as they appear in the text, but spread them across columns and pages if possible.

Table 1. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

Data set	Naive	Flexible	Better?
Breast	95.9± 0.2	96.7± 0.2	√
Cleveland	83.3± 0.6	80.0± 0.6	×
Glass2	61.9± 1.4	83.8± 0.7	√
Credit	74.8± 0.5	78.3± 0.6	
Horse	73.3± 0.9	69.7± 1.0	×
Meta	67.1± 0.6	76.5± 0.5	√
Pima	75.1± 0.6	73.9± 0.5	
Vehicle	44.9± 0.6	61.5± 0.4	√

2.9. Theorems and such

The preferred way is to number definitions, propositions, lemmas, etc. consecutively, within sections, as shown below.

Definition 2.1. A function $f : X \to Y$ is injective if for any $x, y \in X$ different, $f(x) \neq f(y)$.

Using Theorem 2.1 we immediate get the following result:

Proposition 2.2. If f is injective mapping a set X to another set Y, the cardinality of Y is at least as large as that of X

Proof. Left as an exercise to the reader. □

Theorem 2.3 stated next will prove to be useful.

Lemma 2.3. For any $f : X \to Y$ and $g : Y \to Z$ injective functions, $f \circ g$ is injective.

Theorem 2.4. If $f : X \to Y$ is bijective, the cardinality of X and Y are the same.

An easy corollary of Theorem 2.4 is the following:

Corollary 2.5. If $f : X \to Y$ is bijective, the cardinality of X is at least as large as that of Y.

Assumption 2.6. The set X is finite.

Remark 2.7. According to some, it is only the finite case (cf. Theorem 2.6) that is interesting.

2.10. Citations and References

Please use APA reference format regardless of your formatter or word processor. If you rely on the LATEX bibliographic facility, use natbib.sty and icml2025.bst included in the style-file package to obtain this format.

Citations within the text should include the authors' last names and year. If the authors' names are included in the sentence, place only the year in parentheses, for example when referencing Arthur Samuel's pioneering work (?). Otherwise place the entire reference in parentheses with the authors and year separated by a comma (?). List multiple references separated by semicolons (???). Use the 'et al.' construct only for citations with three or more authors or after listing all authors to a publication in an earlier reference (?).

Authors should cite their own work in the third person in the initial version of their paper submitted for blind review. Please refer to Section 2.3 for detailed instructions on how to cite your own papers.

Use an unnumbered first-level section heading for the references, and use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 10 points. The references at the end of this document give examples for journal articles (?), conference publications (?), book chapters (?), books (?), edited volumes (?), technical reports (?), and dissertations (?).

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

Please put some effort into making references complete, presentable, and consistent, e.g. use the actual current name of authors. If using bibtex, please protect capital letters of names and abbreviations in titles, for example, use {B}ayesian or {L}ipschitz in your .bib file.

Accessibility

Authors are kindly asked to make their submissions as accessible as possible for everyone including people with disabilities and sensory or neurological differences. Tips of how to achieve this and what to pay attention to will be provided on the conference website http://icml.cc/.

Software and Data

If a paper is accepted, we strongly encourage the publication of software and data with the camera-ready version of the paper whenever appropriate. This can be done by including a URL in the camera-ready copy. However, do not include URLs that reveal your institution or identity in your submission for review. Instead, provide an anonymous URL or upload the material as "Supplementary Material" into the OpenReview reviewing system. Note that reviewers are not required to look at this material when writing their review.

Acknowledgements

Do not include acknowledgements in the initial version of the paper submitted for blind review.

If a paper is accepted, the final camera-ready version can (and usually should) include acknowledgements. Such acknowledgements should be placed at the end of the section, in an unnumbered section that does not count towards the paper page limit. Typically, this will include thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.

Impact Statement

Authors are required to include a statement of the potential broader impact of their work, including its ethical aspects and future societal consequences. This statement should be in an unnumbered section at the end of the paper (co-located with Acknowledgements – the two may appear in either order, but both must be before References), and does not count toward the paper page limit. In many cases, where the ethical impacts and expected societal implications are those that are well established when advancing the field of Machine Learning, substantial discussion is not required, and a simple statement such as the following will suffice:

"This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here."

The above statement can be used verbatim in such cases, but we encourage authors to think about whether there is content which does warrant further discussion, as this statement will be apparent if the paper is later flagged for ethics review.

A. You can have an appendix here.

You can have as much text here as you want. The main body must be at most 8 pages long. For the final version, one more page can be added. If you want, you can use an appendix like this one.

The \onecolumn command above can be kept in place if you prefer a one-column appendix, or can be removed if you prefer a two-column appendix. Apart from this possible change, the style (font size, spacing, margins, page numbering, etc.) should be kept the same as the main body.