

---

# A Deep Bayesian Inference Architecture

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

The abstract paragraph should be indented 1/2 inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

## 1 Introduction

From a probabilistic perspective, we can view the task of supervised learning as fitting a conditional model of the form  $p(\mathbf{y}|\mathbf{x})$ , which specifies a distribution over outputs given inputs (P44 of Murphy (2022)).

A recently popular approach to unsupervised learning is known as self-supervised learning. In this approach, the task of unsupervised learning is transformed to supervised learning by creating a proxy supervised task. For example, causal language modeling, predict the next word from a word sequence, masked language modeling, mask out words in a sentence and predict them given the surrounding context. The resulting proxy task can also be viewed as a conditional model of the form  $p(\mathbf{y}|\mathbf{x})$ , where both outputs and inputs come from the unlabeled data.

When dealing with high-dimensional data, it is often useful to reduce the dimensionality by projecting it to a lower dimensional subspace which captures the "essence" of the data. (P46)

### 1.1 Bayesian Inference

Bayes theorem is a principle way to calculate a conditional probability.

Bayesian inference is an iterative process utilizing Bayes' Theorem to deduce(infer) a probability distribution based on new observed data coming in iteratively (Harper, 2009). Bayesian inference allows you to update your beliefs iteratively as new information(data) comes in. It works as follows: you have a prior belief about the distribution of your target, then, after you receive some new data, you can update your beliefs by calculating the posterior distribution by Bayes rule. Afterwards, we get even more data come in. So our posterior becomes the new prior. We can update the new prior with the likelihood derived from the new data and again we get a new posterior. This cycle can continue so long as new data comes in, so we can continuously updating our beliefs.

<https://towardsdatascience.com/probability-concepts-explained-bayesian-inference-for-parameter-e>

The posterior is a kind of weighted average (mean) of the prior, where the likelihood are the weights, the evidence are the summation of the weights.

At each Bayesian iteration, the likelihood is provided by the output of the corresponding layer of the neural network.

## 32 1.2 Language Model

33 A goal of language modeling is to learn the joint probability function of sequences of words in a  
34 language (Bengio et al., 2000).

35 <https://www.inference.vc/implicit-bayesian-inference-in-sequence-models/>:

36 We can think of these one-step-ahead predictive distributions as implicitly performing Bayesian  
37 inference.

## 38 1.3 Deep Bayesian Inference Architecture

39 **Advantage** Similar to auxiliary loss, DBI reduce the vanishing gradient problem for earlier lay-  
40 ers, stabilize the training and is used as regularization. [https://stats.stackexchange.com/](https://stats.stackexchange.com/questions/304699/what-is-auxiliary-loss-as-mentioned-in-ppnet-paper)  
41 [questions/304699/what-is-auxiliary-loss-as-mentioned-in-ppnet-paper](https://stats.stackexchange.com/questions/304699/what-is-auxiliary-loss-as-mentioned-in-ppnet-paper)

42 Deep generative models are used to estimate the likelihood of each observation and to create new  
43 samples from the underlying distribution. Suppose the underlying distribution is a categorical  
44 distribution, whose parameters are  $p_i, i \in [1, N]$ . Then the likelihood is defined as  $P(p_i|c)$ , where  $c$   
45 is the context observations. The likelihood works as weights in a Bayesian iteration to update the  
46 prior belief. (Ref: An introduction to deep generative modeling)

## 47 1.4 Related Works

48 **Bayesian Neural Network** A Bayesian neural network (BNN) is commonly defined as a stochastic  
49 artificial neural network trained using Bayesian inference (Jospin et al., 2022). While the prior  
50 distribution in BNN is taken over the possible model parameterization, the prior distribution of Deep  
51 Bayesian Inference is taken over the possible parameterization of data distribution.

## 52 1.5 Retrieval of style files

53 The  $\LaTeX$  style file contains three optional arguments: `final`, which creates a camera-ready copy,  
54 `preprint`, which creates a preprint for submission to, e.g., arXiv, and `nonatbib`, which will not  
55 load the `natbib` package for you in case of package clash.

56 **Preprint option** If you wish to post a preprint of your work online, e.g., on arXiv, using the  
57 NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your  
58 work with the text “Preprint. Work in progress.” in the footer. This version may be distributed as  
59 you see fit. Please **do not** use the `final` option, which should **only** be used for papers accepted to  
60 NeurIPS.

61 At submission time, please omit the `final` and `preprint` options. This will anonymize your  
62 submission and add line numbers to aid review. Please *do not* refer to these line numbers in your  
63 paper as they will be removed during generation of camera-ready copies.

64 The file `neurips_2022.tex` may be used as a “shell” for writing your paper. All you have to do is  
65 replace the author, title, abstract, and text of the paper with your own.

66 The formatting instructions contained in these style files are summarized in Sections 2, 3, and 4  
67 below.

## 68 2 General formatting instructions

69 The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long.  
70 The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points.  
71 Times New Roman is the preferred typeface throughout, and will be selected for you by default.  
72 Paragraphs are separated by  $\frac{1}{2}$  line space (5.5 points), with no indentation.

73 The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal  
74 rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow  $\frac{1}{4}$  inch  
75 space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the  
76 page.

77 For the final version, authors' names are set in boldface, and each name is centered above the  
78 corresponding address. The lead author's name is to be listed first (left-most), and the co-authors'  
79 names (if different address) are set to follow. If there is only one co-author, list both author and  
80 co-author side by side.

81 Please pay special attention to the instructions in Section 4 regarding figures, tables, acknowledgments,  
82 and references.

### 83 **3 Headings: first level**

84 All headings should be lower case (except for first word and proper nouns), flush left, and bold.

85 First-level headings should be in 12-point type.

#### 86 **3.1 Headings: second level**

87 Second-level headings should be in 10-point type.

##### 88 **3.1.1 Headings: third level**

89 Third-level headings should be in 10-point type.

90 **Paragraphs** There is also a `\paragraph` command available, which sets the heading in bold, flush  
91 left, and inline with the text, with the heading followed by 1 em of space.

## 92 **4 Citations, figures, tables, references**

93 These instructions apply to everyone.

### 94 **4.1 Citations within the text**

95 The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as  
96 long as you maintain internal consistency. As to the format of the references themselves, any style is  
97 acceptable as long as it is used consistently.

98 The documentation for `natbib` may be found at

99 `http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

100 Of note is the command `\citet`, which produces citations appropriate for use in inline text. For  
101 example,

102 `\citet{hasselmo}` investigated\dots

103 produces

104 Hasselmo, et al. (1995) investigated...

105 If you wish to load the `natbib` package with options, you may add the following before loading the  
106 `neurips_2022` package:

107 `\PassOptionsToPackage{options}{natbib}`

108 If `natbib` clashes with another package you load, you can add the optional argument `nonatbib`  
109 when loading the style file:

110 `\usepackage[nonatbib]{neurips_2022}`

111 As submission is double blind, refer to your own published work in the third person. That is, use "In  
112 the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers  
113 that are not widely available (e.g., a journal paper under review), use anonymous author names in the  
114 citation, e.g., an author of the form "A. Anonymous."



Figure 1: Sample figure caption.

Table 1: Sample table title

Part		
Name	Description	Size ( $\mu\text{m}$ )
Dendrite	Input terminal	$\sim 100$
Axon	Output terminal	$\sim 10$
Soma	Cell body	up to $10^6$

## 115 4.2 Footnotes

116 Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number<sup>1</sup>  
 117 in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote  
 118 with a horizontal rule of 2 inches (12 picas).

119 Note that footnotes are properly typeset *after* punctuation marks.<sup>2</sup>

## 120 4.3 Figures

121 All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction.  
 122 The figure number and caption always appear after the figure. Place one line space before the figure  
 123 caption and one line space after the figure. The figure caption should be lower case (except for first  
 124 word and proper nouns); figures are numbered consecutively.

125 You may use color figures. However, it is best for the figure captions and the paper body to be legible  
 126 if the paper is printed in either black/white or in color.

## 127 4.4 Tables

128 All tables must be centered, neat, clean and legible. The table number and title always appear before  
 129 the table. See Table 1.

130 Place one line space before the table title, one line space after the table title, and one line space after  
 131 the table. The table title must be lower case (except for first word and proper nouns); tables are  
 132 numbered consecutively.

133 Note that publication-quality tables *do not contain vertical rules*. We strongly suggest the use of the  
 134 booktabs package, which allows for typesetting high-quality, professional tables:

135 <https://www.ctan.org/pkg/booktabs>

136 This package was used to typeset Table 1.

---

<sup>1</sup>Sample of the first footnote.

<sup>2</sup>As in this example.

## 5 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

## 6 Preparing PDF files

Please prepare submission files with paper size “US Letter,” and not, for example, “A4.”

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF file uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdf fonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NeurIPS. Please see <http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf>
- `xfig` “patterned” shapes are implemented with bitmap fonts. Use “solid” shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

```
\usepackage{amsfonts}
```

followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for  $\mathbb{R}$ ,  $\mathbb{N}$  or  $\mathbb{C}$ . You can also use the following workaround for reals, natural and complex:

```
\newcommand{\RR}{\mathbb{R}} %real numbers
\newcommand{\Nat}{\mathbb{N}} %natural numbers
\newcommand{\CC}{\mathbb{C}} %complex numbers
```

Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

### 6.1 Margins in L<sup>A</sup>T<sub>E</sub>X

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the `graphics` bundle documentation (<http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf>)

A number of width problems arise when L<sup>A</sup>T<sub>E</sub>X cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

## References

- Yoshua Bengio, Réjean Ducharme, and Pascal Vincent. 2000. A neural probabilistic language model. *Advances in Neural Information Processing Systems*, 13.
- Marc Harper. 2009. The replicator equation as an inference dynamic. *arXiv preprint arXiv:0911.1763*.
- Laurent Valentin Jospin, Hamid Laga, Farid Boussaid, Wray Buntine, and Mohammed Bennamoun. 2022. Hands-on bayesian neural networks—a tutorial for deep learning users. *IEEE Computational Intelligence Magazine*, 17(2):29–48.

180 Kevin P Murphy. 2022. *Probabilistic machine learning: an introduction*. MIT press.

## 181 Checklist

182 The checklist follows the references. Please read the checklist guidelines carefully for information on  
183 how to answer these questions. For each question, change the default **[TODO]** to **[Yes]** , **[No]** , or  
184 **[N/A]** . You are strongly encouraged to include a **justification to your answer**, either by referencing  
185 the appropriate section of your paper or providing a brief inline description. For example:

- 186 • Did you include the license to the code and datasets? **[Yes]** See Section 2.
- 187 • Did you include the license to the code and datasets? **[No]** The code and the data are  
188 proprietary.
- 189 • Did you include the license to the code and datasets? **[N/A]**

190 Please do not modify the questions and only use the provided macros for your answers. Note that the  
191 Checklist section does not count towards the page limit. In your paper, please delete this instructions  
192 block and only keep the Checklist section heading above along with the questions/answers below.

### 193 1. For all authors...

- 194 (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s  
195 contributions and scope? **[TODO]**
- 196 (b) Did you describe the limitations of your work? **[TODO]**
- 197 (c) Did you discuss any potential negative societal impacts of your work? **[TODO]**
- 198 (d) Have you read the ethics review guidelines and ensured that your paper conforms to  
199 them? **[TODO]**

### 200 2. If you are including theoretical results...

- 201 (a) Did you state the full set of assumptions of all theoretical results? **[TODO]**
- 202 (b) Did you include complete proofs of all theoretical results? **[TODO]**

### 203 3. If you ran experiments...

- 204 (a) Did you include the code, data, and instructions needed to reproduce the main experi-  
205 mental results (either in the supplemental material or as a URL)? **[TODO]**
- 206 (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they  
207 were chosen)? **[TODO]**
- 208 (c) Did you report error bars (e.g., with respect to the random seed after running experi-  
209 ments multiple times)? **[TODO]**
- 210 (d) Did you include the total amount of compute and the type of resources used (e.g., type  
211 of GPUs, internal cluster, or cloud provider)? **[TODO]**

### 212 4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

- 213 (a) If your work uses existing assets, did you cite the creators? **[TODO]**
- 214 (b) Did you mention the license of the assets? **[TODO]**
- 215 (c) Did you include any new assets either in the supplemental material or as a URL?  
216 **[TODO]**
- 217 (d) Did you discuss whether and how consent was obtained from people whose data you’re  
218 using/curating? **[TODO]**
- 219 (e) Did you discuss whether the data you are using/curating contains personally identifiable  
220 information or offensive content? **[TODO]**

### 221 5. If you used crowdsourcing or conducted research with human subjects...

- 222 (a) Did you include the full text of instructions given to participants and screenshots, if  
223 applicable? **[TODO]**
- 224 (b) Did you describe any potential participant risks, with links to Institutional Review  
225 Board (IRB) approvals, if applicable? **[TODO]**
- 226 (c) Did you include the estimated hourly wage paid to participants and the total amount  
227 spent on participant compensation? **[TODO]**